PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, 15/86, C07K 14/82, 14/705, A61K 38/19, C12Q 1/68, C12N 5/10, C12Q 1/00, G01N 33/53

A1

(11) International Publication Number: **WO 97/13855**

(43) International Publication Date: 17 April 1997 (17.04.97)

(21) International Application Number: PCT/EP95/03988

(22) International Filing Date: 10 October 1995 (10.10.95)

(71) Applicant *(for all designated States except US)*: NOVARTIS AG [CH/CH]; Patent and Trademark Dept., Klybeckstrasse 141, CH-4002 Basle (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PLUSCHKE, Gerd [DE/DE]; Am Schönberg 6, D-79249 Merzhausen (DE). SCHMID, Peter [DE/DE]; Goethestrasse 7, D-79540 Lörrach (DE).

(74) Common Representative: NOVARTIS AG; Patent and Trademark Dept., Klybeckstrasse 141, CH-4002 Basle (CH).

(81) Designated States: AL, AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, US, UZ, VN, ARIPO patent (KE, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MELANOMA-ASSOCIATED PROTEIN

(57) Abstract

This invention relates to a melanoma-associated protein (MCSP), a derivative of said protein and to means and methods for the production thereof. The invention is also directed to isolated nucleic acids coding for said melanoma-associated protein, to a method of obtaining such nucleic acid molecules, and to their expression. Furthermore, the invention is directed to uses of said protein and nucleic acid, particularly uses relating to diagnosis, prophylaxis and therapy of tumors producing the protein, such as human malignant melanoma, sarcoma and glioblastoma.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

Melanoma-associated Protein

This invention relates to a melanoma-associated protein (MCSP), a derivative of said protein and to means and methods for the production thereof. The invention is also directed to isolated nucleic acids coding for said melanoma-associated protein, to a method of obtaining such nucleic acid molecules, and to their expression. Furthermore, the invention is directed to uses of said protein and nucleic acid, particularly uses relating to diagnosis, prophylaxis and therapy of tumors producing the protein, such as human malignant melanoma, sarcoma and glioblastoma.

During the last two decades there has been considerable interest in the biology and pathophysiology of human malignant melanoma, in part, because of the poor prognosis and increasing incidence of this disease. The fatal nature of human cutaneous melanoma, which is attributable to poor response to conventional radiation and chemotherapy, has prompted a growing interest in melanoma-associated antigens (MAAs). Such proteins are expressed on melanoma cells but not on normal skin melanocytes and include antigens that are unique for melanoma, or particular stages of melanoma progression, and others that are typical for all tumors of neuroectodermal origin. Based on proven and putative biochemical and immunological characteristics MAAs may be categorized into cell substrate-interacting glycoproteins, ion transport and binding proteins, gangliosides, and receptors for growth factors.

The category of cell substrate-interacting glycoproteins comprises several MAAs of relatively high molecular weight. Up to today, murine monoclonal antibodies (mAb) raised against human melanoma cells or membrane preparations of such cells have been relied upon for identification and partial characterization of these antigens. Thus, human melanoma chondroitin sulfate proteoglycan (MCSP), also referred to as high molecular weight-melanoma associated antigen (HMW-MAA), has been identified with mAb 9.2.27 (Morgan et al., Hybridoma 1, 27-36 (1981)). MCSP is expressed on more than 90% of human melanoma tissues and cultures where 80 to 100 % of cells express MCSP at densities ranging from $1x10^5$ to $6x10^6$ binding sites per cell (Bumol and Reisfeld, Proc. Natl. Acad. Sci. U.S.A. 79, 1245-1249 (1982); Bumol et al., J. Biol. Chem. 267, 12733-12741 (1984); Mueller and Reisfeld in Encyclopedia of Human Biology (Dulbecco, ed.), pp. 957-967, Academic Press, New York (1991)).

MCSP has been reported to be a unique glycoprotein-proteoglycan complex. A 250 kDa molecule is the core glycoprotein of MCSP possessing asparagine-N-linked

oligosaccharides of the high mannose type. Addition of chondroitin sulfate glycosaminolycan polysyccharide chains to serine residues of the core glycoprotein converts the 250 kDa core protein to the high molecular weight proteoglycan form. The molecular mass of the mature proteoglycan containing a full complement of chondroitin sulfate chains has been estimated at 420 to 1000 kDa (Harper and Reisfeld in "Biology of Proteoglycans" (Wight and Mecham eds.), pp. 345-366, Academic Press, Orlando (1987)). The primary structure of the core protein of the rat homologue, referred to as NG2, is known from Nishiyama et al., J. Cell Biology 114, 359-371 (1991).

Proteoglycans have been implicated in growth control, involvement in adhesion, in cell-substratum interaction and cell-cell contacts (Hardingham and Fosang, FASEB J. 6, 861-870 (1992)). Thus, MCSP is found to be expressed on the melanoma cells upper surface on microspikes, consisting of 1-2 µm structures that range up to 20 µm at the cell periphery. These peripheral structures are involved in cell-cell contacts and also form complex footpads that are in contact with the substratum (Bumol et al., 1984, supra; Harper et al., J. Immunol. 132, 2096-2104 (1984); Garrigues et al., J. Natl. Cancer Inst. 71, 259-263 (1986)). Adhesion plaques deposited along the cell membrane also expressed MCSP very well (Harper et al., supra; Harper and Reisfeld, J. Natl. Cancer Inst. 71, 259-263 (1983)). A possible role of this molecule in stabilizing cell-substratum interactions is suggested by the finding that mAb 9.2.27, directed against both the proteoglycan and the core protein, blocks early events of melanoma cell spreading on endothelial basement membranes, while only slightly interfering with cell adhesion. Data indicating that MCSP core protein is expressed on the cell surface in two forms, either modified by the addition of chondroitin sulfate chains or chondroitin sulfate nonmodified, suggest that glycosaminoglycan (GAG) chains may not be necessary for cell surface expression of the core protein. Hence, it seems unlikely that such a modification serves as a marker to segregate molecules on the cell surface (Harper et al., J. Biol. Chem. 261, 3600-3606 (1986)). Furthermore MCSP recognized by mAb 9.2.27 is reported to act as a co-receptor for spreading and focal contact formation in association with $\alpha 4\beta 1$ integrin in melanoma cells, implying a model in which MCSP communicates with $\alpha 4\beta 1$ integrin by an inside-out signaling mechanism (Iida et al., Cancer Research 55, 2177-2185 (1995)).

MCSP also proved to be an effective target for radioimaging of tumors of melanoma patients (Oldham et al., J. Clin. Oncol. 2, 1235-1245 (1995)) and is currently used to target active-specific immunotherapy with antiidiotypic mAb, which bear the internal image of antigenic determinants defined by anti-MCSP mAb (Kusama et al., J. Immunol. 143,

3844-3852 (1989); Chen et al., Cancer Res. 53, 112-119 (1993)). Thus, conjugation of antiidiotypic mAb with a carrier and administration with an adjuvant induced humoral anti-MCSP immunity in about 60% of immunized patients with advanced melanoma (Mittelman et al., Proc. Natl. Acad. Sci. U.S.A. 89, 466-470 (1992)). Development of this anti-MCSP immunity was found to be associated with statistically significant prolongation of survival.

Although some features of MCSP have been described in the literature as shown above, the precise structure of MCSP has not been previously established. In view of the pathological significance of high MCSP expression, particularly in metastatic lesions, there is a need for a better understanding of cellular signal transduction via MCSP as well as in the role of MCSP in the interaction with surrounding cells and tumor spreading. So far, a deeper insight into human malignant melanoma and tumor progression as well as eventual improvements in diagnosis, prophylaxis and therapy of neoplasms showing high MCSP expression has been significantly hampered by the inavailability of MCSP in a purified form and amino acid and nucleic acid sequence information. This lack of knowledge has particularly handicapped the search for human therapeutic agents capable of influencing tumor growth and metastatic spreading.

The present invention has achieved the isolation and sequencing of DNA encoding full-length human MCSP, thus providing the amino acid sequence of human MCSP and enabling the production of MCSP, e.g. by recombinant DNA techniques. Synthesis of a complete cDNA coding for the full-length protein was extemely difficult and could not be achieved by conventional methods. The present invention for the first time enables correlations between MCSP structure and function, thereby providing e.g. means for improved diagnosis, prophylaxis and therapy of a tumor characterized by MCSP expression, e.g. a melanoma, glioma or sarcoma expressing MCSP.

More specifically, the present invention relates to a purified or isolated protein designated MCSP, or a derivative thereof.

As used herein before or hereinafter, the term "purified" or "isolated" is intended to refer to a molecule of the invention in an essentially pure form, said molecule being obtainable from a natural source or by means of genetic engineering. The purified protein, DNA or RNA of the invention may be useful in ways that the protein, DNA and RNA as they naturally occur are not, such as identification of compounds selectively modulating the

expression or the activity of MCSP.

In a preferred embodiment, the invention concerns a protein having the amino acid sequence set forth in SEQ ID NO:2, and particularly a mature protein having the amino acid sequence extending from the amino acid at position 1 (Ala) to the amino acid at position 2293 (Val). Hereinafter, such protein will be referred to as MCSP. The peptide comprising amino acids -29 to -1 of SEQ ID NO:2 represents the MCSP signal peptide. MCSP is found to be an integral membrane protein with a large amino-terminal ectodomain separated from a relatively short cytoplasmic tail by a single hydrophobic transmembrane region.

Included within the scope of " isolated MCSP" or "a protein of the invention", as these terms are understood herein, is any deglycosylated, unglycosylated or glycosylated form of the protein having the amino acid sequence set forth in SEQ ID NO:2, a splice variant encoded by mRNA generated by alternative splicing of a primary MCSP-encoding transcript, and an amino acid mutant of the protein of SEQ ID NO:2.

Additionally, the invention concerns an in vitro generated covalent or aggregative derivative of a protein of the invention.

According to the invention, purified MCSP is essentially free of all naturally occurring substances with which it is typically found in human tissue. For example, MCSP produced by recombinant means will be free of those contaminants typically found in its in vivo physiological milieu. Purified MCSP also encompasses a protein according to the invention in recombinant cell culture.

A beforementioned protein of the invention, or a derivative thereof, displays a biological profile which is qualitatively essentially identical to the profile characteristic of native MCSP, or at least a cross-section of said MCSP- profile. The biological profile in vitro and in vivo includes antigenicity, ligand binding and signal transduction. The biological profile of a protein of the invention, or a particular biological activity thereof, may be evaluated in a suitable assay employing said protein in a purified form or a host cell producing MCSP. In any case, a protein of the invention bears at least one immune epitope in common with MCSP, or mimics such epitope. Such protein is referred to as immunological equivalent of MCSP.

A protein which bears at least one immune epitope in common with MCSP comprises at least eight to about eleven consecutive amino acids of SEQ ID NO:2 and is capable of cross-reacting with an antibody which is specific for native MCSP. Thus, a protein of the invention is capable of competing with native MCSP for binding to an anti-MCSP antibody, e.g. such antibody raised against melanoma cells or a membranous fraction thereof. Examples of such antibodies are mAb 9.2.27 (Bumol and Reisfeld, Proc. Natl. Acad. Sci. U.S.A. 1245-1249 (1982)), mAb 225.28 (European Patent No. 0 380 607, ATCC accession no. HB 10141) and mAb 763.74 (Giacomini et al., J. Immunol. 135, 696 (1985)).

MCSP acts as a cell surface receptor for human type VI collagen. Therefore, an assay suitable for determining the ligand binding activity of a protein of the invention is an assay determining the interaction between said protein of the invention and collagen VI. Such an assay is known in the art (see e.g. Stallcup et al., J. Cell Biol. 111, 3177-3188 (1990); Nishiyama and Stallcup, Mol. Biol. Cell 4, 1097-1108 (1993)) and comprises contacting a cell producing a protein of the invention with collagen VI and assessing the binding to collagen VI to said cell as compared to a suitable negative control, e.g. by immunofluorescent staining. For example, mammalian cells which do not produce endogenous MCSP, or a homologue thereof, but are capable of secreting type VI collagen, such as B28 rat neural cells or U251MG human glioma cells, are transfected with a DNA coding for a membrane-bound protein of the invention. The transfected cells producing said protein of the invention on the cell-surface are assayed for the ability of the protein of the invention to anchor collagen VI to the cell surface. Alternatively, a ligand binding assay may be performed using purified collagen VI, advantageously attached to a solid phase, and an isolated protein of the invention.

Cell surface MCSP is capable of modifying the function and/or activity of $\alpha 4\beta 1$ integrin (Iida et al., J. Cell Biol. 118, 431-444 (1992)). Hence, this ability of a protein of the invention may be tested in a conventional cell adhesion assay (see e.g. Iida et al., J. Cell Biol. 118, 431-444 (1992)) using suitable cells transfected with a DNA encoding a protein of the invention and producing said protein on the cell surface. Signal transduction of a protein of the invention may be studied by evaluating the collaboration between said protein of the invention and $\alpha 4\beta 1$ integrin in the modulation of cell spreading and focal contact adhesion according to methods available in the art, e.g. the method described by Iida et al. (Cancer Research 55, 2177-2185 (1995)).

A glycosylated form of MCSP according to the invention is e.g. MCSP having a native (human) glycosylation pattern, e.g. a MCSP glycoprotein comprising asparagine N-linked oligosaccharides of the high mannose type, or a MCSP proteoglycan containing a partial or full complement of chondroitin sulfate chains, or a glycosylation variant having a glycosylation pattern which is different from that found for native MCSP. A nonglycosylated form of MCSP according to the invention may be obtained by deglycosylation of a glycosylated form of MCSP, e.g. by enzymatic removal of the glycosyl residues, or by expression of a nucleic acid encoding a protein of the invention in suitable prokaryotic cells.

According to the invention, an amino acid mutant (mutein) may be a substitutional, insertional or deletional variant of a protein with the amino acid sequence set forth in SEQ ID NO:2. Contrary to a naturally occurring allelic or interspecies variant, such a mutant is characterized by the predetermined nature of the variation. Substitutions, deletions and insertions may be combined to arrive at an amino acid mutant of the invention.

For example, a substitutional amino acid mutant is any polypeptide having an amino acid sequence substantially identical to the sequence set forth in SEQ ID NO:2, in which one or more residues have been conservatively substituted with a functionally-similar amino acid residue and which is capable of mimicking an MCSP epitope as described herein before. Conservative substitutions include e.g. the substitution of one non-polar (hydrophobic) residue, such as methionine, valine, leucine, isoleucine for another, substitution of one polar (hydrophilic) residue for another, such as between glycine and serine, between arginine and lysine, and between glutamine and asparagine. Substitutional or deletional mutagenesis may be employed to eliminate O- or N-linked glycosylation sites from MCSP. MCSP has 15 potential N-linked glycosylation sites, which are characterized by the occurrence of the acceptor amino acid asparagine (Asn) in the tripepetide sequence Asn-X-Thr(Ser), wherein X can be any of the twenty naturally occurring L-amino acids except possibly aspartic acid (Asp) (Hubbard and Ivatt, Ann. Rev. Biochem. 50, 555-583 (1981)). Potential O-linked glycosylation sites in the MCSP sequence are characterized in that a serine residue precedes a glycine residue. Such Ser/Gly pairs are located at positions 51/52, 178/179, 570/571, 966/967, 1020/1021, 1067/1068, 1131/1132, 1309/1310, 1355/1356, 1475/1476 and 1872/1873 in SEQ ID NO: 2. Deletions of cysteine or other labile amino acid residues may also be desirable, for example to increase the oxidative stability of a protein of the invention.

As defined herein, a deletional amino acid mutant of MCSP also includes a fragment of mature full-length MCSP consisting of eight or more contiguous amino acids, i.e. eight to 2292 contiguous amino acids, of SEQ ID NO:2. According to the invention, such MCSP fragment is a preferred embodiment of a deletional mutant. Preferred are fragments of mature MCSP comprising from about ten to about hundred, particularly, from about ten to about fifty contiguous amino acids of SEQ ID NO:2. A major class of deletional mutants are those involving the transmembrane and/or cytoplasmic region of MCSP. The transmembrane domain succeeds the N-terminal extracellular domain and essentially consists of about 25 amino acids. Extending from about residue 2193 (Met) to about amino acid residue 2217 (Leu) in SEQ ID NO:2, this highly hydrophobic domain has the proper size to span the lipid bilayer of the cellular membrane. The cytoplasmic domain of MCSP follows the transmembrane domain and is the C-terminal sequence of amino acid residues approximately commencing at position 2218 (Arg) in SEQ ID NO:2. Deletion or substitution of either or both of the cytoplasmic and transmembrane domains will facilitate recovery of a recombinant protein of the invention by reducing its cellular or membrane lipid affinity and improving its solubility in water or buffers so that detergents will not be required to maintain the protein in aqueous solution. An example of a deletional mutant involving the transmembrane and the cytoplasmic region of MCSP is the MCSP fragment with the sequence extending from amino acid 1 (Met) to amino acid 1593 (Val) in SEQ ID NO:2. Such deletional mutant and fragments thereof consisting of at least eight, particularly from about ten to about fifty consecutive amino acids of SEQ ID NO:2 are particularly preferred.

Preferred proteins of the invention are mature MCSP having the amino acid sequence set forth in SEQ ID NO:2 in a glycosylated or non-glycosylated form, and a deletional variant thereof, which is a fragment of MCSP as defined above.

A derivative of a protein of the invention is a covalent or aggregative conjugate of said protein with another chemical moiety, said derivative displaying essentially the same biological profile as the underivatized protein of the invention.

An exemplary covalent conjugate according to the invention is a conjugate of a protein of the invention with another protein or peptide, such as a protein comprising a protein of the invention, particularly an MCSP fragment, and a carrier protein suitable for enhancing the in vivo antigenicity of said protein of the invention. A covalent conjugate of the invention further includes a protein of the invention labelled with a detectable group, e.g. a protein

of the invention which is radiolabelled, covalently bound to a rare earth chelate or biotin, or conjugated to a fluorescent moiety.

An aggregative derivative of a protein of the invention is e.g. an adsorption complex of said protein with a cell membrane.

A protein of the invention is obtainable from a natural source, e.g. by isolation from human cells or human tissue expressing MCSP, such as human melanoma tissue, or, preferably, by chemical synthesis or recombinant DNA techniques. Also, a combination of these techniques may be used to obtain a protein of the invention.

Based on the amino acid sequence information provided in SEQ ID NO:2 chemical synthesis of a protein of the invention is performed according to conventional methods known in the art. In general, those methods comprise the sequential addition of one or more amino acid residues to a growing (poly)peptide chain. If required, potentially reactive groups, e.g. free amino or carboxy groups, are protected by a suitable, selectively removable protecting group. Chemical synthesis may be particularly advantageous for fragments of MCSP having no more than about 100, and usually no more than about 20 to 40, amino acid residues.

The invention also provides a method for preparing a protein of the invention, said method being characterized in that suitable host cells producing the protein of the invention are multiplied in vitro or in vivo. Preferably, the host cells are transformed or transfected with a hybrid vector comprising an expression cassette comprising a promoter and a DNA sequence coding for a protein of the invention which DNA is controlled by said promoter. Subsequently, the protein of the invention may be recovered. Recovery comprises e.g. isolating the protein of the invention from the host cells or isolating the host cells comprising the protein, e.g. from the culture broth.

Suitable host cells include eukaryotic cells, e.g. animal cells, plant cells and fungi, and prokaryotic cells, such as gram-positive and gram-negative bacteria, e.g. E. coli.

As used herein, in vitro means ex vivo, thus including e.g. cell culture and tissue culture conditions.

An amino acid mutant, as defined hereinbefore, may be produced e.g. from a DNA

encoding a protein of SEQ ID NO:2, which DNA has been subjected to site-specific in vitro mutagenesis resulting e.g. in an addition, exchange and/or deletion of one or more amino acids. While the site for introducing an amino acid variation is predetermined, the mutation per se need not be predetermined, but random mutagenesis may be performed at the target codon or region. For example, substitutional, deletional and insertional variants are prepared by recombinant methods and screened for immuno-crossreactivity with the native forms of the protein of the invention. Alternatively, mutants of the invention may be prepared by chemical synthesis using methods routinely employed in the art.

A transmembrane and/or cytoplasmic deleted or substituted amino acid mutant of the invention can be produced directly in recombinant cell culture or as a fusion with a signal sequence, preferably a host-homologous signal. For example, in constructing a procaryotic expression vector, the transmembrane and the cytoplasmic domains are deleted in favor of the bacterial alkaline phosphatase, or heat stable enterotoxin II leaders, and for yeast the domains are substituted by yeast invertase, alpha factor or acid phosphatase leaders. In mammalian cell expression the transmembrane and the cytoplasmic domains may be replaced with a mammalian cell viral secretory leader. The advantage of a variant lacking both the transmembrane and the cytoplasmic region is that it is capable of being secreted into the culture medium.

A protein of the invention may also be derivatized in vitro according to conventional methods known in the art.

A protein of the invention, or a derivative thereof, may be used, for example, as immunogen, e.g. to raise MCSP specific immunoreagents, as immunoreagent, in a drug or ligand screening assay, or in a purification method, such as affinity purification of a binding ligand. A protein of the invention, or a derivative thereof, suitable for in vivo administration and capable of competing with endogenous MCSP for an endogenous ligand, e.g. collagen VI, is envisaged as therapeutic agent.

The invention also relates to the use of a protein of the invention, or a derivative thereof, for the generation of a monoclonal or polyclonal antibody, which specifically binds to MCSP. Such anti-MCSP antibody is intended to include immune sera. Particularly useful for this purpose is a MCSP fragment consisting of at least eight or more, preferably eight to about twenty, consecutive amino acids of MCSP of SEQ ID NO:2. The antibodies raised against a protein of the invention may react with a non-glycosylated or glycosylated

form of MCSP, or both.

Monoclonal and particularly polyclonal anti-MCSP antibodies generated against a protein of the invention may be employed as immunoreagents to detect tumor-associated MCSP expression, e.g. in the diagnosis of human malignant melanoma or in the monitoring of melanoma progression and treatment. For example, such antibodies are suitable for MCSP detection in fixed paraffin embedded melanoma lesions. The antibodies are produced in a mammal, e.g. mouse, rat, goat or rabbit, according to methods well-established in the art. For the generation of anti-MCSP antibodies to be used as immunoreagents it is advantageous, if the protein of the invention used as antigen does not comprise a glycosylation site or any part of the transmembrane domain of MCSP. Particularly useful as immunoreagent are antibodies raised against a peptide of the invention which represents a single MCSP determinant.

The invention also relates to the use of a suitably immunogenic protein of the invention, or a suitably immunogenic derivative thereof, as a vaccine, and to a method of vaccinating a human, comprising administration of a suitably immunogenic protein of the invention, or a suitably immunogenic covalent conjugate thereof, to said human. Such a method is intended to also refer to a method of inducing an anti-tumor response in a human comprising administration of a suitably immunogenic protein of the invention, or a suitably immunogenic derivative thereof. A suitably immunogenic protein of the invention, or a suitably immunogenic derivative thereof, is capable of inducing an anti-MCSP response in vivo. A vaccine according to the invention is applicable in the prophylactic and therapeutic treatment of patients having a disposition for or suffering from an MCSP-expressing tumor. As mentioned above, MCSP is a suitable target for active immunotherapy of melanoma, because it is expressed by a high percentage of melanoma lesions involved in metastatic spreading. Thus, a suitably immunogenic protein of the invention, or a derivative thereof, is e.g. a useful agent for the control, treatment or adjuvant treatment of a MCSP-expressing tumor, e.g. melanoma. More specifically, a suitably immunogenic protein of the invention, or a derivative thereof, can be successfully employed e.g. to cause tumor regression and/or prevent tumor recurrence of early stage melanoma patients remaining at risk for metastatic disease after surgery for the primary lesion. Such protein or derivative can be "tailor-made" to bear or mimic a specific determinant of MCSP.

Preferred for use as a vaccine is recombinant MCSP of SEQ ID NO:2, or a fragment thereof consisting of at least eight or more, preferably from eight to about fifty, consecutive amino acids of SEQ ID NO:2. Advantageously, a peptide to be used as vaccine lacks a glycosylation site and

consists of eight to about fourty contiguous amino acids of the extracellular domain of MCSP. Said N-terminal domain consists of about 2192 amino acids and extends from the amino acid residue at position 1 to approximately the amino acid residue at position 2192 in SEQ ID NO:2. The preferred limitation on fragment size is primarily due to the size and purity limitations on synthetic polypeptide imposed by current technologies. Particularly preferred for use as a vaccine are the fragments accentuated above.

Inducement of an appropriate T-cell dependent (memory) response on vivo administration may demand enhancement of the immunogenicity of the protein of the invention, e.g. by conjugation of said protein to a carrier protein, the presence of an adjuvans or expression by vehicles suitable for life vaccination, such as viruses, bacteria or autologous antigen presenting cells. The induction of an anti-MCSP immune response following vaccination may be analyzed according to methods known in the art, e.g. by determination of the anti-MCSP antibody titre in a body fluid of a vaccinated patient, e.g. serum, by means of an enzyme-linked immunoabsorbent-type assay (ELISA).

As carrier protein component, a suitably immunogenic conjugate of the invention may comprise any carrier protein useful in humans, e.g a non-toxic, nonpyrogenic, water soluble, pharmaceutically acceptable carrier protein, preferably such a protein having exposed amino groups. Suitable as a carrier protein component is any proteinaceous molecule containing highly immunogenic promiscuous T-cell epitopes which will bind to a broad range of polymorphic HLA class I and class II gene products, e.g. a microbial, particularly a bacterial, protein, polypeptide or oligopeptide.

Particularly preferred is a covalent conjugate of the invention comprising an above captioned MCSP fragment and a carrier protein component obtainable from a bacterial toxin, e.g. tetanus toxin (see e.g. B. Bizzini in Bacterial Vaccines, Academic Press, 1984: Tetanus, pp. 38-68) or diphteria toxin, which protein component is devoid of toxin activity, but retains the antigenic properties particular to the toxin, e.g. potent immunogenicity, e.g. a mutant diphteria toxin devoid of toxin activity (Uchida et al., J. Pharma Biol. Chem. 218, 3838-3844 (1973)). Such non-toxic carrier protein component is obtainable e.g. by detoxification of the toxin, e.g. in case of tetanus toxin, or by mutation, e.g. in case of diphteria toxin.

Diphteria toxin is obtainable from culture supernatants of Corynebacterium diphteriae PW8 according to the method disclosed by R.K. Holmes, Infect. Immun. 12, 1392 (1975).

The preferred carrier protein to form a covalent conjugate of the invention is the mutant diphteria toxin CRM197. CRM197 is an atoxic protein which crossreacts immunologically with diphteria toxin and is obtainable fromm culture supernatants of C. diphteriae C7 (R.K. Holmes, supra; U.S. Patent No. 4,925,792, which are incorporated herein by reference). The CRM197 protein has the same molecular weight as the diphteric toxin and is composed of a fragment B which is identical as to its function and structure to those of the toxin, and of a fragment A, which is nontoxic and differs from the original fragment by one amino acid.

The carrier protein may be covalently attached to a protein of the invention involving a functional group thereof. The coupling reaction is performed according to methods known in the art in such a way that protein aggregation is avoided. Alternatively, the MCSP-carrier protein conjugate may be produced as a fusion protein by recombinant means.

The invention also relates to an immunogen for use in a mammal comprising a suitably immunogenic protein of the invention, preferably an above specified preferred peptide of the invention, or a derivative thereof. Preferred is such immunogen comprising in a single protein such peptide according to the invention and a carrier protein, as described above.

The invention also concerns pharmaceutical compositions comprising a protein according to the invention. In particular, the invention relates to a pharmaceutical composition comprising an above-specified peptide of the invention in a suitably immunogenic form, e.g. an above-specified preferred peptide of the invention covalently attached to an appropriate carrier protein, as decribed above. The pharmaceutical compositions comprise, for example, a therapeutically effective amount of a protein of the invention in a suitably immunogenic form together or in admixture with pharmaceutically acceptable, inorganic or organic, solid or liquid carriers. Preferred are pharmaceutical compositions additionally comprising an adjuvant, i.e. an agent further increasing the immune response. Possible adjuvants are Freund's complete adjuvant (emulsion of mineral oil, water, and mycobacterial extracts), Freund's incomplete adjuvant (emulsion of water and oil only), mineral gels, e.g. aluminium hydroxide gels, surface active substances such as lysolecithin, polyanions, peptides, BCG (Bacillus Calmette-Guerin), etc.. Particularly preferred are pharmaceutical compositions comprising a suitably immunogenic conjugate of the invention and MF59 (international patent application WO 90/14837) as adjuvant, and, optionally, N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-

[1,2-dipalmitoyl-*sn*-glycero-3-(hydroxyphosphoryloxy]ethylamide (MTP-PE; international patent application WO 90/14837).

Preferred are pharmaceutical compositions for parenteral application. Compositions for intramuscular, subcutaneous or intravenous application are e.g. isotonic aqueous solutions or suspensions, optionally prepared shortly before use from lyophilized or concentrated preparations. The pharmaceutical compositions may be sterilized and contain adjuvants e.g. for conserving, stabilizing, wetting, emulsifying or solubilizing the ingredients, salts for the regulation of the osmotic pressure, buffer and/or compounds regulating the viscosity, e.g. sodium carboxycellulose, dextran, polyvinylpyrrolidine or gelatine. They are prepared by methods known in the art, e.g. by conventional mixing, dissolving or lyophilizing, and contain from approximately 0.01% to approximately 50% of active ingredients. The compositions for injections are processed, filled into ampoules or vials, and sealed under aseptic conditions according to methods known in the art. For example, owing to the solubility in aqueous solutions a conjugate of the invention may be formulated as a "two vial system" with an above adjuvant, e.g. MF59-0.

Preferred is a pharmaceutical composition comprising an above-captioned conjugate of the invention suitable for intramuscular administration in a depot formulation together with an adjuvant.

Also preferred is a pharmaceutical composition comprising a suitably immunogenic protein of the invention, preferably a conjugate of the invention, which is appropriate for mucosal application (H.F. Staats et al., Current Opinion in Immunology 6, 572-583 (1994)), or a stabilized pharmaceutical composition that can be swallowed for oral immunization.

The specific mode of administration and the dosage will be selected by the attending physician taking into account the particulars of the patient, state and type of the disease to be treated, and the like.

Furthermore, a protein of the invention, or a derivative thereof, can be used for the qualitative and quantitative determination of antibodies directed against MCSP. This is especially useful for the detection of an anti-MCSP immune response induced by an anti-melanoma vaccine, such as an antiidiotypic antibody bearing the internal image of MCSP, particularly such antibody disclosed in European Patent Application

EP-A-0 428 485, melanoma cells, or membraneous fractions thereof, melanoma cell lysates, or a suitably immunogenic protein of the invention. For instance, a protein of the invention, or a derivative thereof according to the invention, can be used in any of the known immunoassays which rely on the binding interaction between the idiotopes of the anti-MCSP antibody and said protein of the invention. Examples of such assays are radio-, enzyme, fluorescence, chemiluminescence, immunoprecipitation, latex agglutination, and hemagglutination immunoassays.

This invention also concerns test kits for the qualitative and quantitative determination of antibodies directed against MCSP comprising a protein of the invention and/or derivatives thereof and, optionally, adjuncts.

In a further aspect, the present invention relates to a nucleic acid (DNA, RNA) comprising an isolated, preferably recombinant, nucleic acid (DNA, RNA) coding for a protein of the invention, or a fragment of such a nucleic acid consisting of at least 14 nucleotides.

According to the invention, isolated MCSP-encoding nucleic acid of the invention includes a MCSP-encoding nucleic acid present in other than in the form or setting in which it is found in nature, thus embracing such nucleic acid in ordinarily MCSP expressing cells, where the nucleic acid is in a chromosomal location different from that of natural cells or is otherwise flanked by a different DNA sequence than that found in nature. The MCSP gene maps to human chromosome 15.

In particular, the invention provides a purified or isolated DNA molecule encoding a protein of the invention, or a fragment of such DNA suitable for use as a screening probe as specified hereinafter. By definition, such a DNA comprises a coding single stranded DNA, a double stranded DNA consisting of said coding DNA and complementary DNA thereto, or this complementary (single stranded) DNA itself. Preferred is a DNA coding for an above protein of the invention herein identified as being preferred, or a fragment of said DNA.

Preferred is a DNA coding for the mature protein having the amino acid sequence set forth in SEQ ID NO:1, particularly a DNA having substantially the nucleotide sequence set forth in SEQ ID NO:1, or a DNA coding for a fragment of said protein consisting of at least 14 consecutive amino acids of the amino acid sequence set forth in SEQ ID NO:1 excluding the DNA with the sequence extending from bp 4867 to bp 7898 in SEQ ID

NO:1 and the DNA with the sequence extending from bp 4858 to 5357 in SEQ ID NO:1, respectively. Particularly preferred is a DNA coding for mature MCSP set forth in SEQ ID NO:1, or coding for a MCSP fragment which is accentuated above, e.g a DNA coding for the MCSP fragment extending from amino acid 1 to amino acid 1593 in SEQ ID NO:1, or a portion of said fragment.

It is envisaged that a nucleic acid of the invention can be readily modified by nucleotide substitution, nucleotide deletion, nucleotide insertion or inversion of a nucleotide stretch, and any combination thereof. Such modified sequences can be used to produce a mutein having an amino acid sequence differing from the sequence of MCSP found in nature. Mutagenesis may be predetermined (site-specific) or random. A mutation which is not a silent mutation must not place sequences out of reading frames and preferably will not create complementary regions that could hybridize to produce secondary mRNA structures such as loops or hairpins.

Given the guidance of the present invention, a nucleic acid of the invention is obtainable according to methods well known in the art. The present invention further relates to a process for the preparation of such nucleic acid.

For example, a DNA of the invention is obtainable by chemical synthesis, by recombinant DNA technology or by polymerase chain reaction (PCR). A suitable method for preparing a DNA of the invention may e.g. comprise the synthesis of a number of oligonucleotides, their amplification by PCR methods, and their splicing to give the desired DNA sequence.

Preparation of a DNA of the invention, or a fragment thereof by recombinant DNA technology may involve screening of a suitable cDNA or genomic library. A suitable library is commercially available, e.g. a library employed in the Examples, or can be prepared from human melanoma tissue samples, cell lines and the like. After screening the library, e.g. with a DNA including substantially the entire MCSP coding region or a suitable oligonucleotide (probe) based on a said DNA, positive clones are identified by detecting a hybridization signal; the identified clones are characterized by restriction enzyme mapping and/or DNA sequence analysis, and then examined, e.g. by comparison with the sequences set forth herein, to ascertain whether they include DNA encoding complete MCSP (i.e., if they include translation initiation and termination codons). If the selected clones are incomplete, they may be used to rescreen the same or a different library to obtain overlapping clones. If the library is genomic, then the overlapping clones

may include exons and introns. If the library is a cDNA library, then the overlapping clones will include an open reading frame. In both instances, complete clones may be identified by comparison with the DNA sequences and deduced amino acid sequence provided herein.

In order to detect the presence or any abnormality of endogenous MCSP genetic screening may be carried out using a nucleotide sequence of the invention as hybridization probe. Also, based on the nucleic acid sequences provided herein antisense-type therapeutic agents may be designed.

In addition to being useful for the production of an above mentioned recombinant protein of the invention, a nucleic acid of the invention is useful as probe, thus e.g. enabling those skilled in the art to identify and/or isolate nucleic acid encoding MCSP or a novel non-human homologue thereof. Such probe according to the invention may be unlabeled or labeled with a chemical moiety suitable for ready detection. As a screening probe, there may be employed a DNA or RNA comprising substantially the entire coding region of MCSP, or a suitable oligonucleotide probe based on said DNA. A suitable oligonucleotide probe (for screening involving hybridization) includes a single stranded DNA or RNA that has a sequence of nucleotides that comprises at least 14, preferably at least about 20 to 30, contiguous bases that are the same as (or complementary to) any 14 or more contiguous bases set forth in SEQ ID NO:1. The nucleic acid sequences selected as probes should be of sufficient length and sufficiently unambiguous so that false positive results are minimized. Examplary probes are the oligonucleotides with the sequences set forth in SEQ ID NOs. 14 to 19.

For example, a method suitable for identifying a nucleic acid encoding MCSP, or a novel non-human homologue thereof, comprises contacting a sample comprising MCSP candidate DNA or RNA with a nucleic acid probe described above, and identifying nucleic acid(s) which hybridize (s) to that probe. In particular, nucleic acid according to the invention is useful e.g. in a method for determining the presence of MCSP-mRNA, said method comprising hybridizing the DNA (or RNA) encoding (or complementary to) a protein of the invention, or a fragment of said DNA, to test sample nucleic acid and determining the presence of the desired mRNA, or amplifying, e.g. by PCR, MCSP-RNA using MCSP specific oligonucleotide primers derivable from a nucleic acid sequence provided herein. This method may be employed in tumor diagnosis, e.g. for localization of MCSP mRNA in a tumor, particularly primary melanomas or metastatic lesions of

malignant melanoma. For example, specific in situ hybridization signals for MCSP mRNA expression obtained with antisense RNA probes according to the invention are clearly associated with cells obtained from a metastatic lesion of malignant melanoma. Only non-specific background signals are found with tumor infiltrating cells. Hybridization with sense control RNA yields only non-specific background signals.

The DNA encoding a protein of the invention can be incorporated into vectors for further manipulation. Furthermore, the invention concerns a recombinant DNA which is a hybrid vector comprising at least one of the above mentioned DNAs.

A hybrid vector of the invention comprises an origin of replication or an autonomously replicating sequence, one or more dominant marker sequences and, optionally, expression control sequences, signal sequences and additional restriction sites.

Preferably, the hybrid vector of the invention comprises an above described nucleic acid insert operably linked to an expression control sequence, in particular those described hereinafter.

Vectors typically perform two functions in collaboration with compatible host cells. One function is to facilitate the cloning of the nucleic acid that encodes the protein of the invention, i.e. to produce usable quantities of the nucleic acid (cloning vectors). The other function is to provide for replication and expression of the gene constructs in a suitable host, either by maintenance as an extrachromosomal element or by integration into the host chromosome (expression vectors). A cloning vector comprises the DNAs as described above, an origin of replication or an autonomously replicating sequence, selectable marker sequences, and optionally, signal sequences and additional restriction sites. An expression vector additionally comprises expression control sequences essential for the transcription and translation of the DNA of the invention. Thus an expression vector refers to a recombinant DNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into a suitable host cell, results in expression of the cloned DNA. Suitable expression vectors are well known in the art and include those that are replicable in eukaryotic and/or prokaryotic cells.

Most expression vectors are capable of replication in at least one class of organisms but can be transfected into another organism for expression. For example, a vector is cloned in E. coli and then the same vector is transfected into yeast or mammalian cells even though

it is not capable of replicating independently of the host cell chromosome. DNA may also be amplified by insertion into the host genome. However, the recovery of genomic DNA encoding MCSP is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise MCSP DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, expression and cloning vectors according to the invention contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

Since the amplification of the vectors is conveniently done in E. coli, an E. coli genetic marker and an E. coli origin of replication are advantageously included. These can be obtained from E. coli plasmids, such as pBR322, Bluescript vector or a pUC plasmid.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up MCSP nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes confering resistance to G418 or hygromycin. The mammalian cell transfectants are placed under selection pressure which only those transfectants are uniquely adapted to survive which have taken up and are expressing the marker.

Expression and cloning vectors usually contain a promoter that is recognized by the host organism and is operably linked to MCSP nucleic acid. Such promoter may be inducible or constitutive. The promoter is operably linked to DNA encoding a protein of the invention by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector.

Promoters suitable for use with prokaryotic hosts include, for example, the β-lactamase and lactose promoter systems, alkaline phosphatase, a tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding a

protein of the invention, using linkers or adaptors to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding MCSP.

MCSP gene transcription from vectors in mammalian host cells may be controlled by promoters compatible with the host cell systems, e.g. promoters derived from the genomes of viruses. Suitable plasmids for expression of the protein of the invention in eukaryotic host cells, particularly mammalian cells, are e.g. cytomegalovirus (CMV) promoter-containing vectors, RSV promoter-containing vectors and SV40 promoter-containing vectors and MMTV LTR promoter-containing vectors. Depending on the nature of their regulation, promoters may be constitutive or regulatable by experimental conditions.

Transcription of a DNA encoding a protein according to the invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector.

Construction of vectors according to the invention employs conventional ligation techniques. The various DNA segments of the vector DNA are operatively linked, i.e. they are contiguous and placed into a functional relationship to each other. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a manner known in the art. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing MCSP expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting, e.g. to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), in situ hybridization, using an appropriately labelled probe based on a sequence provided herein, binding assays, immunodetection and functional assays.

The invention further provides host cells capable of producing a protein of the invention and including heterologous (foreign) DNA encoding said protein.

The nucleic acids of the invention can be expressed in a wide variety of host cells, e.g. those mentioned above, that are transformed or transfected with an appropriate expression vector. A protein of the invention may also be expressed as a fusion protein. Recombinant

cells can then be cultured under conditions whereby the protein (s) encoded by the DNA of the invention is (are) expressed.

Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-prositive organisms, such as E. coli, e.g. E. coli K-12 strains, DH5α and HB 101, or Bacilli. Further host cells suitable for MCSP encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. Saccharomyces cerevisiae. Higher eukaryotic cells include insect, amphebian and vertebrate cells, particularly mammalian cells. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. The host cells referred to in this application comprise cells in in vitro culture as well as cells that are within a host animal. Advantageously, a host cell of the invention does not produce endogenous MCSP or a homologue thereof.

Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

Host cells are transfected or transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique, by electroporation or by lipofectin-mediated. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognized when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used. (See, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press).

A DNA of the invention may also be expressed in a non-human transgenic animal, particularly a transgenic warm-blooded animal, and in non-human transgenic tumor cells. Methods for producing a transgenic animal, including mouse, rat, rabbit, sheep and pig, are known in the art and are disclosed, for example, by Hammer et al. (Nature 315,

680-683, (1985)). For instance, an expression unit including a DNA of the invention coding for MCSP together with appropriately positioned expression control sequences, is introduced into pronuclei of fertilized eggs, or in tumor cells. Introduction may be achieved, e.g. by microinjection. Integration of the injected DNA is detected, e.g. by blot analysis of DNA from suitable tissue samples. It is preferred that the DNA be incorporated into the germ line of the animal, so that it is passed to the animal's progeny. Transgenic tumor cells are introduced into a suitable animal.

Furthermore, a knock-out animal may be developed by introducing a mutation in the endogenous MCSP-homologue, thereby generating an animal, which does not express the functional MCSP-homologue gene anymore. For example, in a rat the NG2 gene may be knocked out. A mutated or nonmutated MCSP gene is introduced into the knock-out animal. Expression of human counterpart MCSP on a homologous gene knock-out background has the unique advantage of excluding differences in efficacies of a potential drug on the given protein (in this case MCSP) caused by species-specific sequence differences in said protein.

In a further aspect, the invention relates to an assay for identifying a compound which is capable of interacting with MCSP, comprising contacting cells containing a heterologous DNA encoding a protein of the invention and producing said protein with at least one compound to be tested for its ability to interact with MCSP, and analysing cells for a difference in ligand binding or signal transduction. Suitable analysing methods are known in the art, or may be readily designed based on the known methods and the guidelines provided herein. Preferably, the heterologous DNA comprises substantially the entire coding region. The result obtained in such assay is compared to an assay suitable as a negative control.

Assay methods generally require comparison to various controls. A change in MCSP activity or function is said to be induced by a test compound if such an effect does not occur in the absence of the test compound. An effect of a test compound on a protein of the invention is said to be mediated by said protein if this effect is not observed in cells which do not produce said protein.

For example, by interacting with MCSP compounds may affect melanoma cell growth and spreading, cell-adhesion including cell-substratum interaction and cell-cell contact and MCSP related signal transduction, thus being potential anti-tumor drugs. An assay as

described above is suitable to identify a compound which is capable of inhibiting the binding of collagen VI to MCSP.

The invention particularly relates to the specific embodiments (proteins, nucleic acids, methods for the preparation and uses thereof) as described in the Examples which serve to illustrate the present invention, but should not be construed as limitation thereof.

Example 1: Isolation of MCSP cDNA

A radiolabeled approximately 4.0 kb cDNA encoding the carboxyl-terminus of the rat NG2 transcript (G11, cf. Fig. 2 in Nishiyama et al., J. Cell Biol. 114, 359-371 (1991), is used to screen a λgt11 human melanoma cDNA library prepared from RNA extracted from the M21 human melanoma cell line (Clontech Laboratories, San Francisco, CA) for MCSP candidate cDNA clones. An initial screen of recombinant phages containing dT-primed cDNA yields several NG2 reactive reactive clones.
Recombinant phages ($5x10^5$) are plated at a density of $4x\ 10^4$ plaques per 150 mm petri dish and propagated for 12 hrs at 37°C in the Y 1090r *E. coli* host strain. Phage DNA is transferred to nitrocellulose filters, denatured in 0.5 N NaOH, 1.5 mM NaCl and subsequently neutralized in 0.5 M Tris-HCl, pH 8.0, 1.5 M NaCl. Non-specific nucleic acid binding sites are blocked in a prehybridization medium consisting of 2xSSC (SSC: 150 mM sodium chloride, 15 mM sodium citrate) and 50 µg/ml denatured, sheared salmon sperm DNA at 55°C for 2 hrs. Hybridization reactions are performed under the same conditions for 8 to 12 hrs in the presence of 10 ng/ml of G11 NG2 cDNA fragment (supra) radiolabeled with ^{32}P adCTP to a specific activity of $4x10^8$ cpm/µg by random priming (Sambrook et al., A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)). Following hybridization, the filters are washed in 2xSSC at 55°C and exposed to XAR film (Eastman Kodak) for autoradiography. This screen yields seven NG2 reactive clones, including one clone designated λM3.1, containing 3.1 kb of cDNA. Nucleotide sequence analysis of this isolate by the chain termination method of Sanger et al. (Sambrook et al., supra) indicates an open reading frame encoding 700 carboxyl-terminal amino acid residues with approximately 79% homology to the NG2 protein.

The 3.1 kb isolate of the λM 3.1 clone (supra) is utilized as probe in Northern analysis of polyadenylated RNA extracted from human melanoma cell lines expressing the mAb 9.2.27-reactive antigen. RNA gel blotting and hybridization is done by size-fractionation of 2 µg of polyadenylated RNA on 1.2% formaldehyde-agarose gels and transferred onto

nytran. Prehybridization and hybridization conditions are as described above. Transcript sizes are estimated by comparison with size standards (Gibco-Bethesda Research Laboratories). Hybridization to an 8.0 kb transcript present in M21 melanoma cells (Bumol et al., J. Biol. Chem. 267, 12733-12741 (1991)) and UAC 903 melanoma cells is observed, said transcript being similar in size to the NG2 transcript. The UAC903 cell line expresses approximately five-fold more MCSP transcript than the M21 cell line, an observation that is consistent with a similar elevation in the level of MCSP core protein produced in these cell lines. The cDNA fail to hybridize to RNA from the mAb 9.2.27-unreactive RAJI lymbhoblastoid cell line (Dierich et al., J. Immunol. 112, 1766-1773 (1974)).

A 500 bp fragment from the 5' end of λM3.1 (extending from bp 4858 to 5357 in SEQ ID NO:1) is employed as a probe to screen 5 x 10^5 clones from two independent λgt11 cDNA libraries derived from the human melanoma cell line M21 for overlapping cDNA clones extending further upstream. Exhaustive screening of four independent melanoma cell cDNA libraries with λM3.1 fails to identify clones extending further upstream of the 5' end of λM3.1.
The impossibility to obtain a complete cDNA strand is probably due to MCSP RNA secondary structure. Therefore, a strategy is developed that allows to obtain small overlapping cDNA clones. cDNA synthesis for sequence determination of the entire coding sequence of MCSP core protein is accomplished by polymerase chain reaction (PCR) using a wide range of different primers (see Table 1) and a variety of different RNA denaturation conditions to secure that suitable conditions for each individual portion are present in a sample. Thus, cDNA fragments are obtained which are suitable for PCR amplification.

Experimental details for the generation of PCR-amplified cDNA clones are as follows: RNA is prepared from A375-Met human melanoma cells (Kozlowski et al., J. Natl. Cancer Inst. 72, 913-917 (1984); cell line A 357 is obtainable from the American Type Culture Collection (ATCC) under accession no. ATCC CRL 1619) by the acid guanidinium thiocyanate phenol-chloroform method (Chomzynski and Sacchi, Anal. Biochem. 162, 156 (1987)). Polyadenylated RNA is obtained using a Qiagen Oligotex mRNA preparation kit (Diagen, Hilden, Germany). First strand cDNA is prepared with an M-MuLV Reverse Transcriptase Kit (Life Technologies, Gaithersburg, MD) and either MCSP sequence-specific oligonucleotides, oligo (dT) or random primers in parallel samples. Generally, the best results are obtained with MCSP sequence-specific

oligonucleotides (cf. Table 1, infra). On several occasions, the primary products are reamplified with a nested 5' primer to improve specificity. PCR amplifications are done with the primers indicated in Table 2 applying standard protocols (Rolfs et al., PCR: Clinical Diagnostics and Research, Springer Verlag, Berlin (1992)). Taq, Pwo (Boehringer Mannheim, Germany) or Pfu DNA polymerase (Stratagene, La Jolla, CA) are applied as thermostable polymerases. Anchored PCR is performed after dG tailing of cDNA (Pluschke et al., Eur. J. Immunol. 21, 2749-2754 (1991)). Prior to tailing, first strand cDNA is treated with RNAse H (Life Technologies) and purified with a Glass MAX DNA isolation spin cartridge system (Life Technologies). cDNA is dG-tailed using the Deoxynucleotidyl Terminal Transferase kit (Life Technologies). Oligonucleotides are obtained from Microsynth (Windisch, Switzerland). PCR amplification products of the expected size are isolated from 1% agarose gels and are introduced into plasmids pBluescript KS (Stratagene) or pGEM-1 (Promega, Madison, WI) either by blunt-end cloning into the Hinc II site (Table 1) or by directional cloning (Table 2). Double-stranded plasmid DNAs are sequenced directly with the Sequenase kit (US Biochemical, Cleveland, OH). Sequence data are processed with the aid of a GCG Wisconsin Software Package (Genetics Comuting Group, Madison, WI).

Table 1: Clones used for establishing the MCSP primary nucleotide sequence

Clone	pos.[1)]	Primers for first PCR	for nested PCR
ra25	1-118	P24, SEQ ID NO:3 RA 14, SEQ ID NO:4	not done
ra1	82 - 704	P23, SEQ ID NO:5 RA2, SEQ ID NO:6	P23, SEQ ID NO:5 RA 12, SEQ ID NO:7
ra4	637-1595	P19, SEQ ID NO:8 RA9, SEQ ID NO:9	not done
ra23	1431-2241	RM11, SEQ ID NO:10 RA5, SEQ ID NO:11	not done
an1	2219-3074	P9, SEQ ID NO:12 oligo dC[2)]	P10, SEQ ID NO:13 oligo dC
an38	3013-3466	P7, SEQ ID NO:14 oligo dC[2)]	P6, SEQ ID NO:15 oligo dC
an2	3441-4219	P3, SEQ ID NO:16 oligo dC[2)]	P4, SEQ ID NO:17 oligo dC

an44	4213-4889	BA50, SEQ ID NO:18 oligo dC[2)]	BA51, SEQ ID NO:19 oligo dC
λM3.1	4866-7898	derived from λgt11 cDNA library	

1) "pos." designates length and bp positions of the respective clone with reference to the MCSP nucleotide sequence set forth in SEQ ID NO:1.

2) after dG tailing of cDNA

In a first step, anchored PCR is used to obtain a clone designated an44 (Table 1) , that extends 5' of λM3.1. After dG tailing of cDNA, oligo dC is applied as 5' primer and a sequence corresponding to the 5' end of λM3.1 is used as 3' anti-sense primer in this amplification. Subsequently, a series of seven additional cDNA clones are generated by PCR to cover the entire MCSP coding sequence (see Table 1, infra). Anchored PCR with oligo dC 5' primers are applied for the amplification of three of these overlapping clones (an2, an38 and an1; Table 1), while conventional PCR using 5' sense primers that correspond to rat NG2 sequences are employed for the generation of the remaining four clones (ra23, ra4, ra1 and ra25; Table 1). Taq is used as DNA polymerase. Sequences of the 3' anti-sense primers used for the generation of these PCR clones are complementary to the 5' end of the respective previous clones. In five cases (cf. Table 1), the primary products are reamplified with a nested 5' primer to improve specificity. Nucleotide sequences derived from this first series of PCR clones and from the λgt11 clone are reconfirmed by analyzing a second set of independently derived overlapping PCR clones (Table 2). Discrepancies which are probably caused by mistakes introduced by PCR amplification (Keohavong and Thilly, Proc. Natl. Acad. Sci. USA 86, 9253 (1989)), are resolved by further analyzing independently-derived PCR clones (Table 2).

Table 2: Clones used for reconfirming the MCSP primary sequence

Clone	pos.[1)]	primer
H10/H4	1 - 637	RMP16, SEQ ID NO:20 RMP19, SEQ ID NO:21
G3/G4/ G104	612 - 1197	RMP17, SEQ ID NO:22 RMP20, SEQ ID NO:23
F2/F3	1173 - 2241	RMP11, SEQ ID NO:24 RMP18, SEQ ID NO:25

E9/ E910	2221 - 3316	RMP9, SEQ ID NO:26 RMP10, SEQ ID NO:27
D3/D23	3297 - 4199	RMP8, SEQ ID NO:28 RMP12, SEQ ID NO:29
C2/C20	3757 - 5137	RMP14, SEQ ID NO:30 RMP15, SEQ ID NO:31
B1/B12/ B36	5018 - 5869	RMP3, SEQ ID NO:32 RMP4, SEQ ID NO:33
A1/A11	5848 - 7216	RMP1, SEQ ID NO:34 RMP2, SEQ ID NO:35

[1] "pos." designates length and bp positions of the respective clone with reference to the MCSP nucleotide sequence set forth in SEQ ID NO:1.

The complete coding sequence of the MCSP core protein and the deduced amino acid sequence are shown in SEQ ID NO:1. An open reading frame coding for 2322 amino acids is found. The 3'untranslated region consists of 926 nucleotides. The first 29 amino acids (amino acids -29 to -1; SEQ ID NO:1) represent a putative signal sequence, which is only 48% identical with that of NG2. The subsequent stretch of 18 amino acids is 89% identical. A hydrophobic segment of 25 consecutive amino acid residues near the carboxy terminus (amino acid residues 2193-2217, SEQ ID NO:2) is followed by several basic arginine and lysine residues and thus meets the criteria for a transmembrane domain. Thus, the deduced amino acid sequence of the MCSP core protein predicts an integral membrane protein comprising a large extracellular domain separated from a relatively short cytoplasmic tail (75 amino acids) by a single hydrophobic transmembrane region of 25 amino acids.

The large extracellular domain of MCSP spanning 2192 amino acids can be roughly divided into three structural domains: an amino-terminal domain (amino acids 1-611, SEQ ID NO:1) containing eight cysteines and three serine/glycine pairs; a cysteine-free, a serine/glycine rich domain (amino acids 612-1561, SEQ ID NO:1) including seven such potential attachment sites for glycosaminoglycans; and a third structural domain (amino acids 1562 to 2192, SEQ ID NO:1) with only two cysteines and one serine/glycine pair.

The first structural domain (amino acids 1 to 611), which is approximately 82 % structurally homologous to the corresponding domain in the rat NG2 proteoglycan,

contains three of the 15 potential N-linked glycosylation sites. This domain also appears to have a compact configuration, since it contains eight of the ten cysteines of the entire ectodomain, i.e. four potential disulfide bridges in a region spanned by 611 amino acids.

A key feature of the second structural domain of the MCSP ectodomain (amino acids 612 to 1561; SEQ ID NO:1) is its lack of cysteines in a region spanning 950 amino acids; however, this domain contains seven of the eleven serine/glycine pairs of the MCSP extracellular domain, which can serve as potential chondroitin sulfate attachment sites; however, the signal sequence SerGlyXGly for glycosaminoglycans (GAG) occurs only once (amino acids 1308-1340, SEQ ID NO:1). Six of the 15 potential N-linked MCSP glycosylation sites are found in this domain, which is 79% structurally homologous with its counterpart in the rat NG2 proteoglycan.

The third structural domain of MCSP encompassing 630 amino acids (amino acids 1562 to 2192, SEQ ID NO:1) is approximately 75% homologous in structure with the corresponding domain of NG2. This domain consists of two cysteines, separated by 105 amino acids and likely forms a disulfide bridge. The domain has only one potential GAG attachment site indicated by one serine/glycine pair and contains six of the 15 potential N-linked glycosylation sites of the MCSP ectodomain. This is in contrast to the corresponding NG2 domain that features eight cysteines, one serine/glycine pair and five of its 11 potential N-linked glycosylation sites. The major difference between the deduced sequences of NG2 and MCSP is evident between amino acid residues 2043 and 2091. Thus, for NG2, it is reported that a cluster of six cysteine residues is present in this region (Nishiyama et al., supra, Fig. 3). An alignment of the MCSP and NG2 gene portion encoding for this region reveals three additional bases in the MCSP sequence, which are not found in the NG2 gene. The first additional base in position 6128 (SEQ ID NO:1) causes a difference in the reading frame, which continues after the second additional base in position 6244, but is resolved after the third additional base in position 6273. The three additional bases are found both in the λgt11 (??) cDNA clone and in several independent PCR clones derived from the human melanoma cell lines M21 and A375-Met, respectively.

Example 2: Localization of MCSP-Encoding mRNA in Melanoma Lesions by In Situ Hybridization

Three cDNA fragments corresponding to different regions of the MCSP core protein coding sequence are used as riboprobe templates for in situ hybridization experiments.

Clones 11, ra23 and an44 (cf. Table 1) carry MCSP core protein inserts of 559 bp (3756-4314 in pGEM-1), 811 pb (1431-2241 in pBluescript KS) and 677 bp (42113-4889 in pGEM-1), respectively. Sense and anti-sense RNA probes are labeled according to the instructions of the manufacturer (RNA Transkription Kit, Boehringer) with αS^{35}-UTP (more than 400 Ci/mmol, Amersham) to a specific activity of more than 10^9 dpm/µg. Labeled riboprobes are extracted with phenol/chloroform and free nucleotides are removed by passage over a Sephadex G50 column. RNA is precipitated in 2.2 M ammonium acetate in 77% ethanol overnight at -20°C and is resuspended to an approximate activity of 250,000 cpm/µl (5x concentrated stock solution) in 50% v/v dionized formamide containing 20 mM dithiothreithol and stored at -70°C.

Biopsies of melanoma skin metastases, primary melanomas and benign nevi are fixed in 4% paraformaldehyde in phosphate buffered salt solution (PBS) and then embedded in paraffin. Paraffin sections (8 µm) are placed on 3-aminoprophyltriethoxysilane-treated slides, which bind sections covalently to the glass surface and prevent loss of sections during experimental procedures. Paraffin sections are deparaffinized in xylene and absolute ethanol and air dried. Following rehydration with ethanol solutions of decreasing concentrations, sections are postfixed with 4% paraformaldehyde in PBS for 5 min, rinsed in PBS and water and depurinated for 20 min with 0.2 N HCl at room temperature. These sections are then treated for 30 min with 2xSSC(0.3 M NaCl, 0.03 M Na-citrate, pH 7.0) at 70°C, dehydrated with increasing ethanol solutions and finally air dried. Pre-hybridization is performed at 54°C for 3 hrs in a solution of 50% v/v deionized formamide, 10% w/v dextransulfate, 0.3 M NaCl, 10 mM Tris, 10 mM sodium phosphate pH 6.8, 20 mM dithiothreitol, 0.2xDenhardts reagent, 0.1 mg/ml *Escherichia coli* RNA and 0.5 µM non-radioactive αS-UTP. Hybridization is done overnight in the same solution, supplemented with $5x10^4$ cpm/µl αS^{35}-UTP-labeled RNA probe in a humified chamber at 54°C. Slides are washed in the hybridization solution lacking dextransulfate, RNA and non-radioactive UTP, but containing 50% v/v deionized formamide and 10 mM dithiothreitol at 55°C, two times for 1 hr, and equilibrated for 15 min in a buffer solution consisting of 0.5 M NaCl, 10 mM Tris, 1 mM EDTA, 10 mM dithiothreitol, pH 7.5. Sections are then treated with 50 µg/ml RNase A in equilibration buffer for 30 min at 37°C to remove non-specifically bound probe. This is followed by washing in 2xSSC for 1 hr and then in 0.1xSSC for 1 hr at 37°C. Slides are sequentially dehydrated in 65%, 85% and 95% (v/v) ethanol solutions containing 300 mM ammonium acetate and in absolute ethanol before being air dried. Sections are coated with a 1.2 dilution of Ilford K5 photoemulsion, air dried and exposed for 12 days in a light safe box containing silica gel

at 4°C. The slides are then placed into D19 developer (Kodak), fixed in 30% sodium thiosulfate and stained with Haematoxylin and Eosin. The pattern of hybridization signals on autoradiographed sections is analyzed with a photomicroscope and brightfield/darkfield illuminations.

All three probes of the MCSP coding sequence reveal comparable results. Biopsies from melanoma skin metastases that react strongly with MCSP-specific antibodies mAb 9.2.27 (Morgan et al., Hybridoma 1, 27-36 (1981)) or 763.74 (Giacomini et al., J. Immunol. 135, 696 (1985)) show abundant hybridization signals in cancer cells with all three anti-sense RNA probes and only background hybridization with the control sense RNA probes. Some hybridization is also detected in samples of benign nevi and normal epidermis, which exhibit no abundant staining with MCSP-specific antibodies.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: CIBA-GEIGY AG
(B) STREET: Klybeckstr. 141
(C) CITY: Basel
(E) COUNTRY: SCHWEIZ
(F) POSTAL CODE (ZIP): 4002
(G) TELEPHONE: +41 61 69 11 11
(H) TELEFAX: + 41 61 696 79 76
(I) TELEX: 962 991

(ii) TITLE OF INVENTION: Melanoma-associated Protein

(iii) NUMBER OF SEQUENCES: 35

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..6966
(D) OTHER INFORMATION:/product= "MCSP"

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION:1..87

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:88..6966

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATG CAG TCC GGC CGC GGC CCC CCA CTT CCA GCC CCC GGC CTG GCC TTG
Met Gln Ser Gly Arg Gly Pro Pro Leu Pro Ala Pro Gly Leu Ala Leu
-29             -25                 -20                 -15

GCT TTG ACC CTG ACT ATG TTG GCC AGA CTT GCA TCC GCG GCT TCC TTC
Ala Leu Thr Leu Thr Met Leu Ala Arg Leu Ala Ser Ala Ala Ser Phe
            -10                 -5                  1

TTC GGT GAG AAC CAC CTG GAG GTG CCT GTG GCC ACG GCT CTG ACC GAC
Phe Gly Glu Asn His Leu Glu Val Pro Val Ala Thr Ala Leu Thr Asp
      5                 10                  15

ATA GAC CTG CAG CTG CAG TTC TCC ACG TCC CAG CCC GAA GCC CTC CTT
Ile Asp Leu Gln Leu Gln Phe Ser Thr Ser Gln Pro Glu Ala Leu Leu
 20                 25                  30                  35
```

```
CTC CTG GCA GCA GGC CCA GCT GAC CAC CTC CTG CTG CAG CTC TAC TCT
Leu Leu Ala Ala Gly Pro Ala Asp His Leu Leu Leu Gln Leu Tyr Ser
                 40                  45                  50

GGA CGC CTG CAG GTC AGA CTT GTT CTG GGC CAG GAG GAG CTG AGG CTG
Gly Arg Leu Gln Val Arg Leu Val Leu Gly Gln Glu Glu Leu Arg Leu
             55                  60                  65

CAG ACT CCA GCA GAG ACG CTG CTG AGT GAC TCC ATC CCC CAC ACT GTG
Gln Thr Pro Ala Glu Thr Leu Leu Ser Asp Ser Ile Pro His Thr Val
         70                  75                  80

GTG CTG ACT GTC GTA GAG GGC TGG GCC ACG TTG TCA GTC GAT GGG TTT
Val Leu Thr Val Val Glu Gly Trp Ala Thr Leu Ser Val Asp Gly Phe
     85                  90                  95

CTG AAC GCC TCC TCA GCA GTC CCA GGA GCC CCC CTA GAG GTC CCC TAT
Leu Asn Ala Ser Ser Ala Val Pro Gly Ala Pro Leu Glu Val Pro Tyr
100                 105                 110                 115

GGG CTC TTT GTT GGG GGC ACT GGG ACC CTT GGC CTG CCC TAC CTG AGG
Gly Leu Phe Val Gly Gly Thr Gly Thr Leu Gly Leu Pro Tyr Leu Arg
                120                 125                 130

GGA ACC AGC CGA CCC CTG AGG GGT TGC CTC CAT GCA GCC ACC CTC AAT
Gly Thr Ser Arg Pro Leu Arg Gly Cys Leu His Ala Ala Thr Leu Asn
            135                 140                 145

GGC CGC AGC CTC CTC CGG CCT CTG ACC CCC GAT GTG CAT GAG GGC TGT
Gly Arg Ser Leu Leu Arg Pro Leu Thr Pro Asp Val His Glu Gly Cys
        150                 155                 160

GCT GAA GAG TTT TCT GCC AGT GAT GAT GTG GCC CTG GGC TTC TCT GGG
Ala Glu Glu Phe Ser Ala Ser Asp Asp Val Ala Leu Gly Phe Ser Gly
    165                 170                 175
```

```
CCC CAC TCT CTG GCT GCC TTC CCT GCC TGG GGC ACT CAG GAC GAA GGA
Pro His Ser Leu Ala Ala Phe Pro Ala Trp Gly Thr Gln Asp Glu Gly
180                 185                 190                 195

ACC CTA GAG TTT ACA CTC ACC ACA CAG AGC CGG CAG GCA CCC TTG GCC
Thr Leu Glu Phe Thr Leu Thr Thr Gln Ser Arg Gln Ala Pro Leu Ala
                200                 205                 210

TTC CAG GCA GGG GGC CGG CGT GGG GAC TTC ATC TAT GTG GAC ATA TTT
Phe Gln Ala Gly Gly Arg Arg Gly Asp Phe Ile Tyr Val Asp Ile Phe
            215                 220                 225

GAG GGC CAC CTG CGG GCC GTG GTG GAG AAG GGC CAG GGT ACC GTA TTG
Glu Gly His Leu Arg Ala Val Val Glu Lys Gly Gln Gly Thr Val Leu
        230                 235                 240

CTC CAC AAC AGT GTG CCT GTG GCC GAT GGG CAG CCC CAT GAG GTC AGT
Leu His Asn Ser Val Pro Val Ala Asp Gly Gln Pro His Glu Val Ser
    245                 250                 255

GTC CAC ATC AAT GCT CAC CGG CTG GAA ATC TCC GTG GAC CAG TAC CCT
Val His Ile Asn Ala His Arg Leu Glu Ile Ser Val Asp Gln Tyr Pro
260                 265                 270                 275

ACG CAT ACT TCG AAC CGA GGA GTC CTC AGC TAC CTG GAG CCA CGG GGC
Thr His Thr Ser Asn Arg Gly Val Leu Ser Tyr Leu Glu Pro Arg Gly
                280                 285                 290

AGT CTC CTT CTC GGG GGG CTG GAT GCA GAG GCC TCT CGT CAC CTC CAG
Ser Leu Leu Leu Gly Gly Leu Asp Ala Glu Ala Ser Arg His Leu Gln
            295                 300                 305

GAA CAC CGC CTG GGC CTG ACA CCA GAG GCC ACC AAT GCC TCC CTG CTG
Glu His Arg Leu Gly Leu Thr Pro Glu Ala Thr Asn Ala Ser Leu Leu
        310                 315                 320
```

```
GGC TGC ATG GAA GAC CTC AGT GTC AAT GGC CAG AGG CGG GGG CTG CGG      1
Gly Cys Met Glu Asp Leu Ser Val Asn Gly Gln Arg Arg Gly Leu Arg
    325                 330                 335

GAA GCT TTG CTG ACG CGC AAC ATG GCA GCC GGC TGC AGG CTG GAG GAG      1
Glu Ala Leu Leu Thr Arg Asn Met Ala Ala Gly Cys Arg Leu Glu Glu
340                 345                 350                 355

GAG GAG TAT GAG GAC GAT GCC TAT GGC CAT TAT GAA GCT TTC TCC ACC      1
Glu Glu Tyr Glu Asp Asp Ala Tyr Gly His Tyr Glu Ala Phe Ser Thr
                360                 365                 370

CTG GCT CCC GAG GCT TGG CCA GCC ATG GAG CTG CCT GAG CCA TGC GTG      1
Leu Ala Pro Glu Ala Trp Pro Ala Met Glu Leu Pro Glu Pro Cys Val
            375                 380                 385

CCT GAG CCA GGG CTG CCT CCT GTC TTT GCC AAT TTC ACC CAG CTG CTG      1
Pro Glu Pro Gly Leu Pro Pro Val Phe Ala Asn Phe Thr Gln Leu Leu
        390                 395                 400

ACT ATC AGC CCA CTG GTG GTG GCC GAG GGT GGC ACA GCC TGG CTT GAG      1
Thr Ile Ser Pro Leu Val Val Ala Glu Gly Gly Thr Ala Trp Leu Glu
    405                 410                 415

TGG AGG CAT GTG CAG CCC ACG CTG GAC CTG ATG GAG GCT GAG CTG CGC      1
Trp Arg His Val Gln Pro Thr Leu Asp Leu Met Glu Ala Glu Leu Arg
420                 425                 430                 435

AAA TCC CAG GTG CTG TTC AGC GTG ACC CGA GGG GCA CAC TAT GGC GAG      1
Lys Ser Gln Val Leu Phe Ser Val Thr Arg Gly Ala His Tyr Gly Glu
                440                 445                 450

CTC GAG CTG GAC ATC CTG GGT GCC CAG GCA CGA AAA ATG TTC ACC CTC      1
Leu Glu Leu Asp Ile Leu Gly Ala Gln Ala Arg Lys Met Phe Thr Leu
            455                 460                 465
```

```
CTG GAC GTG GTG AAC CGC AAG GCC CGC TTC ATC CAC GAT GGC TCT GAG      1
Leu Asp Val Val Asn Arg Lys Ala Arg Phe Ile His Asp Gly Ser Glu
        470                 475                 480

GAC ACC TCC GAC CAG CTG GTG CTG GAG GTG TCG GTG ACG GCT CGG GTG      1
Asp Thr Ser Asp Gln Leu Val Leu Glu Val Ser Val Thr Ala Arg Val
    485                 490                 495

CCC ATG CCC TCA TGC CTT CGG AGG GGC CAA ACA TAC CTC CTG CCC ATC      1
Pro Met Pro Ser Cys Leu Arg Arg Gly Gln Thr Tyr Leu Leu Pro Ile
500                 505                 510                 515

CAG GTC AAC CCT GTC AAT GAC CCA CCC CAC ATC ATC TTC CCA CAT GGC      1
Gln Val Asn Pro Val Asn Asp Pro Pro His Ile Ile Phe Pro His Gly
                520                 525                 530

AGC CTC ATG GTG ATC CTG GAA CAC ACG CAG AAG CCG CTG GGG CCT GAG      1
Ser Leu Met Val Ile Leu Glu His Thr Gln Lys Pro Leu Gly Pro Glu
            535                 540                 545

GTT TTC CAG GCC TAT GAC CCG GAC TCT GCC TGT GAG GGC CTC ACC TTC      1
Val Phe Gln Ala Tyr Asp Pro Asp Ser Ala Cys Glu Gly Leu Thr Phe
        550                 555                 560

CAG GTC CTT GGC ACC TCC TCT GGC CTC CCC GTG GAG CGC CGA GAC CAG      1
Gln Val Leu Gly Thr Ser Ser Gly Leu Pro Val Glu Arg Arg Asp Gln
    565                 570                 575

CCT GGG GAG CCG GCG ACC GAG TTC TCC TGC CGG GAG TTG GAG GCC GGC      1
Pro Gly Glu Pro Ala Thr Glu Phe Ser Cys Arg Glu Leu Glu Ala Gly
580                 585                 590                 595

AGC CTA GTC TAT GTC CAC TGC GGT GGT CCT GCA CAG GAC TTG ACG TTC      1
Ser Leu Val Tyr Val His Cys Gly Gly Pro Ala Gln Asp Leu Thr Phe
                600                 605                 610
```

```
CGG GTC AGC GAT GGA CTG CAG GCC AGC CCC CCG GCC ACG CTG AAG GTG
Arg Val Ser Asp Gly Leu Gln Ala Ser Pro Pro Ala Thr Leu Lys Val
            615                 620                 625

GTG GCC ATC CGG CCG GCC ATA CAG ATC CAC CGC AGC ACA GGG TTG CGA
Val Ala Ile Arg Pro Ala Ile Gln Ile His Arg Ser Thr Gly Leu Arg
        630                 635                 640

CTG GCC CAA GGC TCT GCC ATG CCC ATC TTG CCC GCC AAC CTG TCG GTG
Leu Ala Gln Gly Ser Ala Met Pro Ile Leu Pro Ala Asn Leu Ser Val
    645                 650                 655

GAG ACC AAT GCC GTG GGG CAG GAT GTG AGC GTG CTG TTC CGC GTC ACT
Glu Thr Asn Ala Val Gly Gln Asp Val Ser Val Leu Phe Arg Val Thr
660                 665                 670                 675

GGG GCC CTG CAG TTT GGG GAG CTG CAG AAG CAT AGT ACA GGT GGG GTG
Gly Ala Leu Gln Phe Gly Glu Leu Gln Lys His Ser Thr Gly Gly Val
                680                 685                 690

GAG GGT GCT GAG TGG TGG GCC ACA CAG GCG TTC CAC CAG CGG GAT GTG
Glu Gly Ala Glu Trp Trp Ala Thr Gln Ala Phe His Gln Arg Asp Val
            695                 700                 705

GAG CAG GGC CGC GTG AGG TAC CTG AGC ACT GAC CCA CAG CAC CAC GCT
Glu Gln Gly Arg Val Arg Tyr Leu Ser Thr Asp Pro Gln His His Ala
        710                 715                 720

TAC GAC ACC GTG GAG AAC CTG GCC CTG GAG GTG CAG GTG GGC CAG GAG
Tyr Asp Thr Val Glu Asn Leu Ala Leu Glu Val Gln Val Gly Gln Glu
    725                 730                 735

ATC CTG AGC AAT CTG TCC TTC CCA GTG ACC ATC CAG AGA GCC ACT GTG
Ile Leu Ser Asn Leu Ser Phe Pro Val Thr Ile Gln Arg Ala Thr Val
740                 745                 750                 755
```

```
TGG ATG CTG CGG CTG GAG CCA CTG CAC ACT CAG AAC ACC CAG CAG GAG
Trp Met Leu Arg Leu Glu Pro Leu His Thr Gln Asn Thr Gln Gln Glu
                760                 765                 770

ACC CTC ACC ACA GCC CAC CTG GAG GCC ACC CTG GAG GAG GCA GGC CCA
Thr Leu Thr Thr Ala His Leu Glu Ala Thr Leu Glu Glu Ala Gly Pro
            775                 780                 785

AGC CCC CCA ACC TTC CAT TAT GAG GTG GTT CAG GCT CCC AGG AAA GGC
Ser Pro Pro Thr Phe His Tyr Glu Val Val Gln Ala Pro Arg Lys Gly
        790                 795                 800

AAC CTT CAA CTA CAG GGC ACA AGG CTG TCA GAT GGC CAG GGC TTC ACC
Asn Leu Gln Leu Gln Gly Thr Arg Leu Ser Asp Gly Gln Gly Phe Thr
    805                 810                 815

CAG GAT GAC ATA CAG GCT GGC CGG GTG ACC TAT GGG GCC ACA GCT CGT
Gln Asp Asp Ile Gln Ala Gly Arg Val Thr Tyr Gly Ala Thr Ala Arg
820                 825                 830                 835

GCC TCA GAG GCA GTC GAG GAC ACC TTC CGT TTC CGT GTC ACA GCT CCA
Ala Ser Glu Ala Val Glu Asp Thr Phe Arg Phe Arg Val Thr Ala Pro
                840                 845                 850

CCA TAT TTC TCC CCA CTC TAT ACC TTC CCC ATC CAC ATT GGT GGT GAC
Pro Tyr Phe Ser Pro Leu Tyr Thr Phe Pro Ile His Ile Gly Gly Asp
            855                 860                 865

CCA GAT GCG CCT GTC CTC ACC AAT GTC CTC CTC GTG GTG CCT GAG GGT
Pro Asp Ala Pro Val Leu Thr Asn Val Leu Leu Val Val Pro Glu Gly
        870                 875                 880

GGT GAG GGT GTC CTC TCT GCT GAC CAC CTC TTT GTC AAG AGT CTC AAC
Gly Glu Gly Val Leu Ser Ala Asp His Leu Phe Val Lys Ser Leu Asn
    885                 890                 895
```

```
AGT GCC AGC TAC CTC TAT GAG GTC ATG GAG CGG CCC CGC CTT GGG AGG
Ser Ala Ser Tyr Leu Tyr Glu Val Met Glu Arg Pro Arg Leu Gly Arg
900                 905                 910                 915

TTG GCT TGG CGT GGG ACA CAG GAC AAG ACC ACT ATG GTG ACA TCC TTC
Leu Ala Trp Arg Gly Thr Gln Asp Lys Thr Thr Met Val Thr Ser Phe
                920                 925                 930

ACC AAT GAA GAC CTG TTG CGT GGC CGG CTG GTC TAC CAG CAT GAT GAC
Thr Asn Glu Asp Leu Leu Arg Gly Arg Leu Val Tyr Gln His Asp Asp
            935                 940                 945

TCC GAG ACC ACA GAA GAT GAT ATC CCA TTT GTT GCT ACC CGC CAG GGC
Ser Glu Thr Thr Glu Asp Asp Ile Pro Phe Val Ala Thr Arg Gln Gly
        950                 955                 960

GAG AGC AGT GGT GAC ATG GCC TGG GAG GAG GTA CGG GGT GTC TTC CGA
Glu Ser Ser Gly Asp Met Ala Trp Glu Glu Val Arg Gly Val Phe Arg
    965                 970                 975

GTG GCC ATC CAG CCC GTG AAT GAC CAC GCC CCT GTG CAG ACC ATC AGC
Val Ala Ile Gln Pro Val Asn Asp His Ala Pro Val Gln Thr Ile Ser
980                 985                 990                 995

CGG ATC TTC CAT GTG GCC CGG GGT GGG CGG CGG CTG CTG ACT ACA GAC
Arg Ile Phe His Val Ala Arg Gly Gly Arg Arg Leu Leu Thr Thr Asp
                1000                1005                1010

GAC GTG GCC TTC AGC GAT GCT GAC TCG GGC TTT GCT GAC GCC CAG CTG
Asp Val Ala Phe Ser Asp Ala Asp Ser Gly Phe Ala Asp Ala Gln Leu
            1015                1020                1025

GTG CTT ACC CGC AAG GAC CTC CTC TTT GGC AGT ATC GTG GCC GTA GAT
Val Leu Thr Arg Lys Asp Leu Leu Phe Gly Ser Ile Val Ala Val Asp
        1030                1035                1040
```

```
GAG CCC ACG CGG CCC ATC TAC CGC TTC ACC CAG GAG GAC CTC AGG AAG
Glu Pro Thr Arg Pro Ile Tyr Arg Phe Thr Gln Glu Asp Leu Arg Lys
    1045                1050                1055

AGG CGA GTA CTG TTC GTG CAC TCA GGG GCT GAC CGT GGC TGG ATC CAG
Arg Arg Val Leu Phe Val His Ser Gly Ala Asp Arg Gly Trp Ile Gln
1060                1065                1070                1075

CTG CAG GTG TCC GAC GGG CAA CAC CAG GCC ACT GCG CTG CTG GAG GTG
Leu Gln Val Ser Asp Gly Gln His Gln Ala Thr Ala Leu Leu Glu Val
                1080                1085                1090

CAG GCC TCG GAA CCC TAC CTC CGT GTG GCC AAC GGC TCC AGC CTT GTG
Gln Ala Ser Glu Pro Tyr Leu Arg Val Ala Asn Gly Ser Ser Leu Val
            1095                1100                1105

GTC CCT CAA GGA GGC CAG GGC ACC ATC GAC ACG GCC GTG CTC CAC CTG
Val Pro Gln Gly Gly Gln Gly Thr Ile Asp Thr Ala Val Leu His Leu
        1110                1115                1120

GAC ACC AAC CTC GAC ATC CGC AGT GGG GAT GAG GTC CAC TAC CAC GTC
Asp Thr Asn Leu Asp Ile Arg Ser Gly Asp Glu Val His Tyr His Val
    1125                1130                1135

ACA GCT GGC CCT CGC TGG GGA CAA CTA GTC CGG GCT GGT CAG CCA GCC
Thr Ala Gly Pro Arg Trp Gly Gln Leu Val Arg Ala Gly Gln Pro Ala
1140                1145                1150                1155

ACA GCC TTC TCC CAG CAG GAC CTG CTG GAT GGG GCC GTT CTC TAT AGC
Thr Ala Phe Ser Gln Gln Asp Leu Leu Asp Gly Ala Val Leu Tyr Ser
                1160                1165                1170

CAC AAT GGC AGC CTC AGC CCC GAA GAC ACC ATG GCC TTC TCC GTG GAA
His Asn Gly Ser Leu Ser Pro Glu Asp Thr Met Ala Phe Ser Val Glu
            1175                1180                1185
```

GCA GGG CCA GTG CAC ACG GAT GCC ACC CTA CAA GTG ACC ATT GCC CTA
Ala Gly Pro Val His Thr Asp Ala Thr Leu Gln Val Thr Ile Ala Leu
1190 1195 1200

GAG GGC CCA CTG GCC CCA CTG AAG CTG GTC CGG CAC AAG AAG ATC TAC
Glu Gly Pro Leu Ala Pro Leu Lys Leu Val Arg His Lys Lys Ile Tyr
1205 1210 1215

GTC TTC CAG GGA GAG GCA GCT GAG ATC AGA AGG GAC CAG CTG GAG GCA
Val Phe Gln Gly Glu Ala Ala Glu Ile Arg Arg Asp Gln Leu Glu Ala
1220 1225 1230 1235

GCC CAG GAG GCA GTG CCA CCT GCA GAC ATC GTA TTC TCA GTG AAG AGC
Ala Gln Glu Ala Val Pro Pro Ala Asp Ile Val Phe Ser Val Lys Ser
1240 1245 1250

CCA CCG AGT GCC GGC TAC CTG GTG ATG GTG TCG CGT GGC GCC TTG GCA
Pro Pro Ser Ala Gly Tyr Leu Val Met Val Ser Arg Gly Ala Leu Ala
1255 1260 1265

GAT GAG CCA CCC AGC CTG GAC CCT GTG CAG AGC TTC TCC CAG GAG GCA
Asp Glu Pro Pro Ser Leu Asp Pro Val Gln Ser Phe Ser Gln Glu Ala
1270 1275 1280

GTG GAC ACA GGC AGG GTC CTG TAC CTG CAC TCC CGC CCT GAG GCC TGG
Val Asp Thr Gly Arg Val Leu Tyr Leu His Ser Arg Pro Glu Ala Trp
1285 1290 1295

AGC GAT GCC TTC TCG CTG GAT GTG GCC TCA GGC CTG GGT GCT CCC CTC
Ser Asp Ala Phe Ser Leu Asp Val Ala Ser Gly Leu Gly Ala Pro Leu
1300 1305 1310 1315

GAG GGC GTC CTT GTG GAG CTG GAG GTG CTG CCC GCT GCC ATC CCA CTA
Glu Gly Val Leu Val Glu Leu Glu Val Leu Pro Ala Ala Ile Pro Leu
1320 1325 1330

GAG GCG CAA AAC TTC AGC GTC CCT GAG GGT GGC AGC CTC ACC CTG GCC
Glu Ala Gln Asn Phe Ser Val Pro Glu Gly Gly Ser Leu Thr Leu Ala
1335 1340 1345

CCT CCA CTG CTC CGT GTC TCC GGG CCC TAC TTC CCC ACT CTC CTG GGC
Pro Pro Leu Leu Arg Val Ser Gly Pro Tyr Phe Pro Thr Leu Leu Gly
1350 1355 1360

CTC AGC CTG CAG GTG CTG GAG CCA CCC CAG CAT GGA CCC CTG CAG AAG
Leu Ser Leu Gln Val Leu Glu Pro Pro Gln His Gly Pro Leu Gln Lys
1365 1370 1375

GAG GAC GGA CCT CAA GCC AGG ACC CTC AGC GCC TTC TCC TGG AGA ATG
Glu Asp Gly Pro Gln Ala Arg Thr Leu Ser Ala Phe Ser Trp Arg Met
1380 1385 1390 1395

GTG GAA GAG CAG CTG ATC CGC TAC GTG CAT GAC GGG AGC GAG ACA CTG
Val Glu Glu Gln Leu Ile Arg Tyr Val His Asp Gly Ser Glu Thr Leu
1400 1405 1410

ACA GAC AGT TTT GTC CTG ATG GCT AAT GCC TCC GAG ATG GAT CGC CAG
Thr Asp Ser Phe Val Leu Met Ala Asn Ala Ser Glu Met Asp Arg Gln
1415 1420 1425

AGC CAT CCT GTG GCC TTC ACT GTC ACT GTC CTG CCT GTC AAT GAC CAA
Ser His Pro Val Ala Phe Thr Val Thr Val Leu Pro Val Asn Asp Gln
1430 1435 1440

CCC CCC ATC CTC ACT ACA AAC ACA GGC CTG CAG ATG TGG GAG GGG GCC
Pro Pro Ile Leu Thr Thr Asn Thr Gly Leu Gln Met Trp Glu Gly Ala
1445 1450 1455

ACT GCG CCC ATC CCT GCG GAG GCT CTG AGG AGC ACG GAC GGC GAC TCT
Thr Ala Pro Ile Pro Ala Glu Ala Leu Arg Ser Thr Asp Gly Asp Ser
1460 1465 1470 1475

```
GGG TCT GAG GAT CTG GTC TAC ACC ATC GAG CAG CCC AGC AAC GGG CGG
Gly Ser Glu Asp Leu Val Tyr Thr Ile Glu Gln Pro Ser Asn Gly Arg
                1480                1485                1490

GTA GTG CTG CGG GGG GCG CCG GGC ACT GAG GTG CGC AGC TTC ACG CAG
Val Val Leu Arg Gly Ala Pro Gly Thr Glu Val Arg Ser Phe Thr Gln
            1495                1500                1505

GCC CAG CTG GAC GGC GGG CTC GTG CTG TTC TCA CAC AGA GGA ACC CTG
Ala Gln Leu Asp Gly Gly Leu Val Leu Phe Ser His Arg Gly Thr Leu
        1510                1515                1520

GAT GGA GGC TTC CCG TTC CGC CTC TCT GAC GGC GAG CAC ACT TCC CCC
Asp Gly Gly Phe Pro Phe Arg Leu Ser Asp Gly Glu His Thr Ser Pro
    1525                1530                1535

GGA CAC TTC TTC CGA GTG ACG GCC CAG AAG CAA GTG CTC CTC TCG CTG
Gly His Phe Phe Arg Val Thr Ala Gln Lys Gln Val Leu Leu Ser Leu
1540                1545                1550                1555

AAG GGC AGC CAG ACA CTG ACT GTC TGC CCA GGG TCC GTC CAG CCA CTC
Lys Gly Ser Gln Thr Leu Thr Val Cys Pro Gly Ser Val Gln Pro Leu
                1560                1565                1570

AGC AGT CAG ACC CTC AGG GCC AGC TCC AGC GCA GGC ACT GAC CCC CAG
Ser Ser Gln Thr Leu Arg Ala Ser Ser Ser Ala Gly Thr Asp Pro Gln
            1575                1580                1585

CTC CTG CTC TAC CGT GTG GTG CGG GGC CCC CAG CTA GGC CGG CTG TTC
Leu Leu Leu Tyr Arg Val Val Arg Gly Pro Gln Leu Gly Arg Leu Phe
        1590                1595                1600

CAC GCC CAG CAG GAC AGC ACA GGG GAG GCC CTG GTG AAC TTC ACT CAG
His Ala Gln Gln Asp Ser Thr Gly Glu Ala Leu Val Asn Phe Thr Gln
    1605                1610                1615
```

GCA GAG GTC TAC GCT GGG AAT ATT CTG TAT GAG CAT GAG ATG CCC CCC
Ala Glu Val Tyr Ala Gly Asn Ile Leu Tyr Glu His Glu Met Pro Pro
1620 1625 1630 1635

GAG CCC TTT TGG GAG GCC CAT GAT ACC CTA GAG CTC CAG CTG TCC TCG
Glu Pro Phe Trp Glu Ala His Asp Thr Leu Glu Leu Gln Leu Ser Ser
1640 1645 1650

CCG CCT GCC CGG GAC GTG GCC GCC ACC CTT GCT GTG GCT GTG TCT TTT
Pro Pro Ala Arg Asp Val Ala Ala Thr Leu Ala Val Ala Val Ser Phe
1655 1660 1665

GAG GCT GCC TGT CCC CAG CGC CCC AGC CAC CTC TGG AAG AAC AAA GGT
Glu Ala Ala Cys Pro Gln Arg Pro Ser His Leu Trp Lys Asn Lys Gly
1670 1675 1680

CTC TGG GTC CCC GAG GGC CAG CGG GCC AGG ATC ACC GTG GCT GCT CTG
Leu Trp Val Pro Glu Gly Gln Arg Ala Arg Ile Thr Val Ala Ala Leu
1685 1690 1695

GAT GCC TCC AAT CTC TTG GCC AGC GTT CCA TCA CCC CAG CGC TCA GAG
Asp Ala Ser Asn Leu Leu Ala Ser Val Pro Ser Pro Gln Arg Ser Glu
1700 1705 1710 1715

CAT GAT GTG CTC TTC CAG GTC ACA CAG TTC CCC AGC CGG GGC CAG CTG
His Asp Val Leu Phe Gln Val Thr Gln Phe Pro Ser Arg Gly Gln Leu
1720 1725 1730

TTG GTG TCC GAG GAG CCC CTC CAT GCT GGG CAG CCC CAC TTC CTG CAG
Leu Val Ser Glu Glu Pro Leu His Ala Gly Gln Pro His Phe Leu Gln
1735 1740 1745

TCC CAG CTG GCT GCA GGG CAG CTA GTG TAT GCC CAC GGC GGT GGG GGC
Ser Gln Leu Ala Ala Gly Gln Leu Val Tyr Ala His Gly Gly Gly Gly
1750 1755 1760

```
ACC CAG CAG GAT GGC TTC CAC TTT CGT GCC CAC CTC CAG GGG CCA GCA
Thr Gln Gln Asp Gly Phe His Phe Arg Ala His Leu Gln Gly Pro Ala
    1765                1770                1775

GGG GCC TCC GTG GCT GGA CCC CAA ACC TCA GAG GCC TTT GCC ATC ACG
Gly Ala Ser Val Ala Gly Pro Gln Thr Ser Glu Ala Phe Ala Ile Thr
1780                1785                1790                1795

GTG AGG GAT GTA AAT GAG CGG CCC CCT CAG CCA CAG GCC TCT GTC CCA
Val Arg Asp Val Asn Glu Arg Pro Pro Gln Pro Gln Ala Ser Val Pro
                1800                1805                1810

CTC CGG CTC ACC CGA GGC TCT CGT GCC CCC ATC TCC CGG GCC CAG CTG
Leu Arg Leu Thr Arg Gly Ser Arg Ala Pro Ile Ser Arg Ala Gln Leu
            1815                1820                1825

AGT GTG GTG GAC CCA GAC TCA GCT CCT GGG GAG ATT GAG TAC GAG GTC
Ser Val Val Asp Pro Asp Ser Ala Pro Gly Glu Ile Glu Tyr Glu Val
        1830                1835                1840

CAG CGG GCA CCC CAC AAC GGC TTC CTC AGC CTG GTG GGT GGT GGC CTG
Gln Arg Ala Pro His Asn Gly Phe Leu Ser Leu Val Gly Gly Gly Leu
    1845                1850                1855

GGG CCC GTG ACC CGC TTC ACG CAA GCC GAT GTG GAT TCA GGG CGG CTG
Gly Pro Val Thr Arg Phe Thr Gln Ala Asp Val Asp Ser Gly Arg Leu
1860                1865                1870                1875

GCC TTC GTG GCC AAC GGG AGC AGC GTG GCA GGC ATC TTC CAG CTG AGC
Ala Phe Val Ala Asn Gly Ser Ser Val Ala Gly Ile Phe Gln Leu Ser
                1880                1885                1890

ATG TCT GAT GGG GCC AGC CCA CCC CTG CCC ATG TCC CTG GCT GTG GAC
Met Ser Asp Gly Ala Ser Pro Pro Leu Pro Met Ser Leu Ala Val Asp
            1895                1900                1905
```

```
ATC CTA CCA TCC GCC ATC GAG GTG CAG CTG CGG GCA CCC CTG GAG GTG
Ile Leu Pro Ser Ala Ile Glu Val Gln Leu Arg Ala Pro Leu Glu Val
        1910                1915                1920

CCC CAA GCT TTG GGG CGC TCC TCA CTG AGC CAG CAG CAG CTC CGG GTG
Pro Gln Ala Leu Gly Arg Ser Ser Leu Ser Gln Gln Gln Leu Arg Val
    1925                1930                1935

GTT TCA GAT CGG GAG GAG CCA GAG GCA GCA TAC CGC CTC ATC CAG GGA
Val Ser Asp Arg Glu Glu Pro Glu Ala Ala Tyr Arg Leu Ile Gln Gly
1940                1945                1950                1955

CCC CAG TAT GGG CAT CTC CTG GTG GGC GGG CGG CCC ACC TCG GCC TTC
Pro Gln Tyr Gly His Leu Leu Val Gly Gly Arg Pro Thr Ser Ala Phe
                1960                1965                1970

AGC CAA TTC CAG ATA GAC CAG GGC GAG GTG GTC TTT GCC TTC ACC AAC
Ser Gln Phe Gln Ile Asp Gln Gly Glu Val Val Phe Ala Phe Thr Asn
            1975                1980                1985

TTC TCC TCC TCT CAT GAC CAC TTC AGA GTC CTG GCA CTG GCT AGG GGT
Phe Ser Ser Ser His Asp His Phe Arg Val Leu Ala Leu Ala Arg Gly
        1990                1995                2000

GTC AAT GCA TCA GCC GTA GTG AAC GTC ACT GTG AGG GCT CTG CTG CAT
Val Asn Ala Ser Ala Val Val Asn Val Thr Val Arg Ala Leu Leu His
    2005                2010                2015

GTG TGG GCA GGT GGG CCA TGG CCC CAG GGT GCC ACC CTG CGC CTG GAC
Val Trp Ala Gly Gly Pro Trp Pro Gln Gly Ala Thr Leu Arg Leu Asp
2020                2025                2030                2035

CCC ACC GTC CTA GAT GCT GGC GAG CTG GCC AAC CGC ACA GGC AGT GTG
Pro Thr Val Leu Asp Ala Gly Glu Leu Ala Asn Arg Thr Gly Ser Val
                2040                2045                2050
```

CCG CGC TTC CGC CTC CTG GAG GGA CCC CGG CAT GGC CGC GTG GTC CGC
Pro Arg Phe Arg Leu Leu Glu Gly Pro Arg His Gly Arg Val Val Arg
2055 2060 2065

GTG CCC CGA GCC AGG ACG GAG CCC GGG GGC AGC CAG CTG GTG GAG CAG
Val Pro Arg Ala Arg Thr Glu Pro Gly Gly Ser Gln Leu Val Glu Gln
2070 2075 2080

TTC ACT CAG CAG GAC CTT GAG GAC GGG AGG CTG GGG CTG GAG GTG GGC
Phe Thr Gln Gln Asp Leu Glu Asp Gly Arg Leu Gly Leu Glu Val Gly
2085 2090 2095

AGG CCA GAG GGG AGG GCC CCC GGC CCC GCA GGT GAC AGT CTC ACT CTG
Arg Pro Glu Gly Arg Ala Pro Gly Pro Ala Gly Asp Ser Leu Thr Leu
2100 2105 2110 2115

GAG CTG TGG GCA CAG GGC GTC CCG CCT GCT GTG GCC TCC CTG GAC TTT
Glu Leu Trp Ala Gln Gly Val Pro Pro Ala Val Ala Ser Leu Asp Phe
2120 2125 2130

GCC ACT GAG CCT TAC AAT GCT GCC CGG CCC TAC AGC GTG GCC CTG CTC
Ala Thr Glu Pro Tyr Asn Ala Ala Arg Pro Tyr Ser Val Ala Leu Leu
2135 2140 2145

AGT GTC CCC GAG GCC GCC CGG ACG GAA GCA GGG AAG CCA GAG AGC AGC
Ser Val Pro Glu Ala Ala Arg Thr Glu Ala Gly Lys Pro Glu Ser Ser
2150 2155 2160

ACC CCC ACA GGC GAG CCA GGC CCC ATG GCA TCC AGC CCT GAG CCC GCT
Thr Pro Thr Gly Glu Pro Gly Pro Met Ala Ser Ser Pro Glu Pro Ala
2165 2170 2175

GTG GCC AAG GGA GGC TTC CTG AGC TTC CTT GAG GCC AAC ATG TTC AGC
Val Ala Lys Gly Gly Phe Leu Ser Phe Leu Glu Ala Asn Met Phe Ser
2180 2185 2190 2195

GTC ATC ATC CCC ATG TGC CTG GTA CTT CTG CTC CTG GCG CTC ATC CTG
Val Ile Ile Pro Met Cys Leu Val Leu Leu Leu Ala Leu Ile Leu
2200 2205 2210

CCC CTG CTC TTC TAC CTC CGA AAA CGC AAC AAG ACG GGC AAG CAT GAC
Pro Leu Leu Phe Tyr Leu Arg Lys Arg Asn Lys Thr Gly Lys His Asp
2215 2220 2225

GTC CAG GTC CTG ACT GCC AAG CCC CGC AAC GGC CTG GCT GGT GAC ACC
Val Gln Val Leu Thr Ala Lys Pro Arg Asn Gly Leu Ala Gly Asp Thr
2230 2235 2240

GAG ACC TTT CGC AAG GTG GAG CCA GGC CAG GCC ATC CCG CTC ACA GCT
Glu Thr Phe Arg Lys Val Glu Pro Gly Gln Ala Ile Pro Leu Thr Ala
2245 2250 2255

GTG CCT GGC CAG GGG CCC CCT CCA GGA GGC CAG CCT GAC CCA GAG CTG
Val Pro Gly Gln Gly Pro Pro Pro Gly Gly Gln Pro Asp Pro Glu Leu
2260 2265 2270 2275

CTG CAG TTC TGC CGG ACA CCC AAC CCT GCC CTT AAG AAT GGC CAG TAC
Leu Gln Phe Cys Arg Thr Pro Asn Pro Ala Leu Lys Asn Gly Gln Tyr
2280 2285 2290

TGG GTG TGAAGGCCTG GCCTGGGCCC AGATGCTGAT CGGGCCAGGG ACAGGCTTGC
Trp Val

CCATGTCCCG GGCCCCATTG CTTCCATGCC CGGTGCTGTC TGAGTATCCC CAGAGCAAGA 707

GAGACCTGGA GACACCAGGG GTGGAGGGTC CTGGGAGATA GTCCCAGGGG TCCGGGACAG 713

AGTGGAGTCA AGAGCTGGAA CCTCCCTCAG CTCACTCCGA GCCTGGAGAA CTGCAGGGGC 719

CAAGGTGGAG GCAGGCTTAA GTTCAGTCCT CCTGCCCTGG AGCTGGTTTG GGCTGTCAAA 725

ACCAGGGTAA CCTCCTACAT GGGTCATGAC TCTGGGTCCT GGGTCTGTGA CCTTGGGTAA 731

GTCGCGCCTG ACCCAGGCTG CTAAGAGGGC AAGGAGAAGG AAGTACCCTG GGGAGGGAAG 737

GGACAGAGGA AGCTATTCCT GGCTTTTCTA CTCCAACCCA GGCCACCCTT TGTCTCTNCC 743

CCAGAGTTGA GAAAAAAACT TCCTCCCCTG GTTTTTTAGG GAGATGGTAT CCCCTGGAGT 749

AGAGGGCAAG AGGAGAGAGC GCCTCCAGTC TAGAAGGCAT AAGCCAATAG GATAATATAT 755

TCAGGGTGCA GGGTGGGTAG GTTGCTCTGG GGATGGGTTT ATTTAAGGGA GATTGCAAGG 761

AAGCTATTTA ACATGGTGCT GAGCTAGCCA GGACTGATGG AGCCCCTGGG GGTGTGGGAT 767

GGAGGAGGGT CTGCAGCCAG TTCATTCCCA GGGCCCCATC TTGATGGGCC AAGGGCTAAA 773

CATGCATGTG TCAGTGGCTT TGGAGCAGGC TAGGCTGGGG CTCATCGAGG GTCTCAGGCC 779

GAGGCCACTG TAGTGCCAGT GCCCCCCTGA GGACTAGGGC AGGCAGCTGG GGGCACTTGG 785

TTCCATGGAG CCTGGATAAA CAGTGCTTTG GAGGCTCTGG AAAAAAAAAA AAAAAAAAAA 791

AA 793

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2322 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Ser Gly Arg Gly Pro Pro Leu Pro Ala Pro Gly Leu Ala Leu
-29 -25 -20 -15

Ala Leu Thr Leu Thr Met Leu Ala Arg Leu Ala Ser Ala Ala Ser Phe
-10 -5 1

Phe Gly Glu Asn His Leu Glu Val Pro Val Ala Thr Ala Leu Thr Asp
5 10 15

Ile Asp Leu Gln Leu Gln Phe Ser Thr Ser Gln Pro Glu Ala Leu Leu
20 25 30 35

Leu Leu Ala Ala Gly Pro Ala Asp His Leu Leu Leu Gln Leu Tyr Ser
40 45 50

Gly Arg Leu Gln Val Arg Leu Val Leu Gly Gln Glu Glu Leu Arg Leu
55 60 65

Gln Thr Pro Ala Glu Thr Leu Leu Ser Asp Ser Ile Pro His Thr Val
70 75 80

Val Leu Thr Val Val Glu Gly Trp Ala Thr Leu Ser Val Asp Gly Phe
85 90 95

Leu Asn Ala Ser Ser Ala Val Pro Gly Ala Pro Leu Glu Val Pro Tyr
100 105 110 115

Gly Leu Phe Val Gly Gly Thr Gly Thr Leu Gly Leu Pro Tyr Leu Arg
120 125 130

Gly Thr Ser Arg Pro Leu Arg Gly Cys Leu His Ala Ala Thr Leu Asn
135 140 145

Gly Arg Ser Leu Leu Arg Pro Leu Thr Pro Asp Val His Glu Gly Cys
150 155 160

Ala Glu Glu Phe Ser Ala Ser Asp Asp Val Ala Leu Gly Phe Ser Gly
165 170 175

Pro His Ser Leu Ala Ala Phe Pro Ala Trp Gly Thr Gln Asp Glu Gly
180 185 190 195

Thr Leu Glu Phe Thr Leu Thr Thr Gln Ser Arg Gln Ala Pro Leu Ala
200 205 210

Phe Gln Ala Gly Gly Arg Arg Gly Asp Phe Ile Tyr Val Asp Ile Phe
215 220 225

Glu Gly His Leu Arg Ala Val Val Glu Lys Gly Gln Gly Thr Val Leu
230 235 240

Leu His Asn Ser Val Pro Val Ala Asp Gly Gln Pro His Glu Val Ser
245 250 255

Val His Ile Asn Ala His Arg Leu Glu Ile Ser Val Asp Gln Tyr Pro
260 265 270 275

Thr His Thr Ser Asn Arg Gly Val Leu Ser Tyr Leu Glu Pro Arg Gly
280 285 290

Ser Leu Leu Leu Gly Gly Leu Asp Ala Glu Ala Ser Arg His Leu Gln
295 300 305

Glu His Arg Leu Gly Leu Thr Pro Glu Ala Thr Asn Ala Ser Leu Leu
310 315 320

Gly Cys Met Glu Asp Leu Ser Val Asn Gly Gln Arg Arg Gly Leu Arg
325 330 335

Glu Ala Leu Leu Thr Arg Asn Met Ala Ala Gly Cys Arg Leu Glu Glu
340 345 350 355

Glu Glu Tyr Glu Asp Asp Ala Tyr Gly His Tyr Glu Ala Phe Ser Thr
360 365 370

Leu Ala Pro Glu Ala Trp Pro Ala Met Glu Leu Pro Glu Pro Cys Val
375 380 385

Pro Glu Pro Gly Leu Pro Pro Val Phe Ala Asn Phe Thr Gln Leu Leu
390 395 400

Thr Ile Ser Pro Leu Val Val Ala Glu Gly Gly Thr Ala Trp Leu Glu
405 410 415

Trp Arg His Val Gln Pro Thr Leu Asp Leu Met Glu Ala Glu Leu Arg
420 425 430 435

Lys Ser Gln Val Leu Phe Ser Val Thr Arg Gly Ala His Tyr Gly Glu
440 445 450

Leu Glu Leu Asp Ile Leu Gly Ala Gln Ala Arg Lys Met Phe Thr Leu
455 460 465

Leu Asp Val Val Asn Arg Lys Ala Arg Phe Ile His Asp Gly Ser Glu
470 475 480

Asp Thr Ser Asp Gln Leu Val Leu Glu Val Ser Val Thr Ala Arg Val
485 490 495

Pro Met Pro Ser Cys Leu Arg Arg Gly Gln Thr Tyr Leu Leu Pro Ile
500 505 510 515

Gln Val Asn Pro Val Asn Asp Pro Pro His Ile Ile Phe Pro His Gly
520 525 530

Ser Leu Met Val Ile Leu Glu His Thr Gln Lys Pro Leu Gly Pro Glu
535 540 545

Val Phe Gln Ala Tyr Asp Pro Asp Ser Ala Cys Glu Gly Leu Thr Phe
550 555 560

Gln Val Leu Gly Thr Ser Ser Gly Leu Pro Val Glu Arg Arg Asp Gln
565 570 575

Pro Gly Glu Pro Ala Thr Glu Phe Ser Cys Arg Glu Leu Glu Ala Gly
580 585 590 595

Ser Leu Val Tyr Val His Cys Gly Gly Pro Ala Gln Asp Leu Thr Phe
600 605 610

Arg Val Ser Asp Gly Leu Gln Ala Ser Pro Pro Ala Thr Leu Lys Val
615 620 625

Val Ala Ile Arg Pro Ala Ile Gln Ile His Arg Ser Thr Gly Leu Arg
630 635 640

Leu Ala Gln Gly Ser Ala Met Pro Ile Leu Pro Ala Asn Leu Ser Val
645 650 655

Glu Thr Asn Ala Val Gly Gln Asp Val Ser Val Leu Phe Arg Val Thr
660 665 670 675

Gly Ala Leu Gln Phe Gly Glu Leu Gln Lys His Ser Thr Gly Gly Val
680 685 690

Glu Gly Ala Glu Trp Trp Ala Thr Gln Ala Phe His Gln Arg Asp Val
695 700 705

Glu Gln Gly Arg Val Arg Tyr Leu Ser Thr Asp Pro Gln His His Ala
710 715 720

Tyr Asp Thr Val Glu Asn Leu Ala Leu Glu Val Gln Val Gly Gln Glu
725 730 735

Ile Leu Ser Asn Leu Ser Phe Pro Val Thr Ile Gln Arg Ala Thr Val
740 745 750 755

Trp Met Leu Arg Leu Glu Pro Leu His Thr Gln Asn Thr Gln Gln Glu
760 765 770

Thr Leu Thr Thr Ala His Leu Glu Ala Thr Leu Glu Glu Ala Gly Pro
775 780 785

Ser Pro Pro Thr Phe His Tyr Glu Val Val Gln Ala Pro Arg Lys Gly
790 795 800

Asn Leu Gln Leu Gln Gly Thr Arg Leu Ser Asp Gly Gln Gly Phe Thr
805 810 815

Gln Asp Asp Ile Gln Ala Gly Arg Val Thr Tyr Gly Ala Thr Ala Arg
820 825 830 835

Ala Ser Glu Ala Val Glu Asp Thr Phe Arg Phe Arg Val Thr Ala Pro
840 845 850

Pro Tyr Phe Ser Pro Leu Tyr Thr Phe Pro Ile His Ile Gly Gly Asp
855 860 865

Pro Asp Ala Pro Val Leu Thr Asn Val Leu Leu Val Val Pro Glu Gly
870 875 880

Gly Glu Gly Val Leu Ser Ala Asp His Leu Phe Val Lys Ser Leu Asn
885 890 895

Ser Ala Ser Tyr Leu Tyr Glu Val Met Glu Arg Pro Arg Leu Gly Arg
900 905 910 915

Leu Ala Trp Arg Gly Thr Gln Asp Lys Thr Thr Met Val Thr Ser Phe
920 925 930

Thr Asn Glu Asp Leu Leu Arg Gly Arg Leu Val Tyr Gln His Asp Asp
935 940 945

Ser Glu Thr Thr Glu Asp Asp Ile Pro Phe Val Ala Thr Arg Gln Gly
950 955 960

Glu Ser Ser Gly Asp Met Ala Trp Glu Glu Val Arg Gly Val Phe Arg
965 970 975

Val Ala Ile Gln Pro Val Asn Asp His Ala Pro Val Gln Thr Ile Ser
980 985 990 995

Arg Ile Phe His Val Ala Arg Gly Gly Arg Arg Leu Leu Thr Thr Asp
1000 1005 1010

Asp Val Ala Phe Ser Asp Ala Asp Ser Gly Phe Ala Asp Ala Gln Leu
1015 1020 1025

Val Leu Thr Arg Lys Asp Leu Leu Phe Gly Ser Ile Val Ala Val Asp
1030 1035 1040

Glu Pro Thr Arg Pro Ile Tyr Arg Phe Thr Gln Glu Asp Leu Arg Lys
1045 1050 1055

Arg Arg Val Leu Phe Val His Ser Gly Ala Asp Arg Gly Trp Ile Gln
1060 1065 1070 1075

Leu Gln Val Ser Asp Gly Gln His Gln Ala Thr Ala Leu Leu Glu Val
1080 1085 1090

Gln Ala Ser Glu Pro Tyr Leu Arg Val Ala Asn Gly Ser Ser Leu Val
1095 1100 1105

Val Pro Gln Gly Gly Gln Gly Thr Ile Asp Thr Ala Val Leu His Leu
1110 1115 1120

Asp Thr Asn Leu Asp Ile Arg Ser Gly Asp Glu Val His Tyr His Val
1125 1130 1135

Thr Ala Gly Pro Arg Trp Gly Gln Leu Val Arg Ala Gly Gln Pro Ala
1140 1145 1150 1155

Thr Ala Phe Ser Gln Gln Asp Leu Leu Asp Gly Ala Val Leu Tyr Ser
1160 1165 1170

His Asn Gly Ser Leu Ser Pro Glu Asp Thr Met Ala Phe Ser Val Glu
1175 1180 1185

Ala Gly Pro Val His Thr Asp Ala Thr Leu Gln Val Thr Ile Ala Leu
1190 1195 1200

Glu Gly Pro Leu Ala Pro Leu Lys Leu Val Arg His Lys Lys Ile Tyr
1205 1210 1215

Val Phe Gln Gly Glu Ala Ala Glu Ile Arg Arg Asp Gln Leu Glu Ala
1220 1225 1230 1235

Ala Gln Glu Ala Val Pro Pro Ala Asp Ile Val Phe Ser Val Lys Ser
1240 1245 1250

Pro Pro Ser Ala Gly Tyr Leu Val Met Val Ser Arg Gly Ala Leu Ala
1255 1260 1265

Asp Glu Pro Pro Ser Leu Asp Pro Val Gln Ser Phe Ser Gln Glu Ala
1270 1275 1280

Val Asp Thr Gly Arg Val Leu Tyr Leu His Ser Arg Pro Glu Ala Trp
1285 1290 1295

Ser Asp Ala Phe Ser Leu Asp Val Ala Ser Gly Leu Gly Ala Pro Leu
1300 1305 1310 1315

Glu Gly Val Leu Val Glu Leu Glu Val Leu Pro Ala Ala Ile Pro Leu
1320 1325 1330

Glu Ala Gln Asn Phe Ser Val Pro Glu Gly Gly Ser Leu Thr Leu Ala
1335 1340 1345

Pro Pro Leu Leu Arg Val Ser Gly Pro Tyr Phe Pro Thr Leu Leu Gly
1350 1355 1360

Leu Ser Leu Gln Val Leu Glu Pro Pro Gln His Gly Pro Leu Gln Lys
1365 1370 1375

Glu Asp Gly Pro Gln Ala Arg Thr Leu Ser Ala Phe Ser Trp Arg Met
1380 1385 1390 1395

Val Glu Glu Gln Leu Ile Arg Tyr Val His Asp Gly Ser Glu Thr Leu
1400 1405 1410

Thr Asp Ser Phe Val Leu Met Ala Asn Ala Ser Glu Met Asp Arg Gln
1415 1420 1425

Ser His Pro Val Ala Phe Thr Val Thr Val Leu Pro Val Asn Asp Gln
1430 1435 1440

Pro Pro Ile Leu Thr Thr Asn Thr Gly Leu Gln Met Trp Glu Gly Ala
1445 1450 1455

Thr Ala Pro Ile Pro Ala Glu Ala Leu Arg Ser Thr Asp Gly Asp Ser
1460 1465 1470 1475

Gly Ser Glu Asp Leu Val Tyr Thr Ile Glu Gln Pro Ser Asn Gly Arg
1480 1485 1490

Val Val Leu Arg Gly Ala Pro Gly Thr Glu Val Arg Ser Phe Thr Gln
1495 1500 1505

Ala Gln Leu Asp Gly Gly Leu Val Leu Phe Ser His Arg Gly Thr Leu
1510 1515 1520

Asp Gly Gly Phe Pro Phe Arg Leu Ser Asp Gly Glu His Thr Ser Pro
1525 1530 1535

Gly His Phe Phe Arg Val Thr Ala Gln Lys Gln Val Leu Leu Ser Leu
1540 1545 1550 1555

Lys Gly Ser Gln Thr Leu Thr Val Cys Pro Gly Ser Val Gln Pro Leu
1560 1565 1570

Ser Ser Gln Thr Leu Arg Ala Ser Ser Ser Ala Gly Thr Asp Pro Gln
1575 1580 1585

Leu Leu Leu Tyr Arg Val Val Arg Gly Pro Gln Leu Gly Arg Leu Phe
1590 1595 1600

His Ala Gln Gln Asp Ser Thr Gly Glu Ala Leu Val Asn Phe Thr Gln
1605 1610 1615

Ala Glu Val Tyr Ala Gly Asn Ile Leu Tyr Glu His Glu Met Pro Pro
1620 1625 1630 1635

Glu Pro Phe Trp Glu Ala His Asp Thr Leu Glu Leu Gln Leu Ser Ser
1640 1645 1650

Pro Pro Ala Arg Asp Val Ala Ala Thr Leu Ala Val Ala Val Ser Phe
1655 1660 1665

Glu Ala Ala Cys Pro Gln Arg Pro Ser His Leu Trp Lys Asn Lys Gly
1670 1675 1680

Leu Trp Val Pro Glu Gly Gln Arg Ala Arg Ile Thr Val Ala Ala Leu
1685 1690 1695

Asp Ala Ser Asn Leu Leu Ala Ser Val Pro Ser Pro Gln Arg Ser Glu
1700 1705 1710 1715

His Asp Val Leu Phe Gln Val Thr Gln Phe Pro Ser Arg Gly Gln Leu
1720 1725 1730

Leu Val Ser Glu Glu Pro Leu His Ala Gly Gln Pro His Phe Leu Gln
1735 1740 1745

Ser Gln Leu Ala Ala Gly Gln Leu Val Tyr Ala His Gly Gly Gly Gly
1750 1755 1760

Thr Gln Gln Asp Gly Phe His Phe Arg Ala His Leu Gln Gly Pro Ala
1765 1770 1775

Gly Ala Ser Val Ala Gly Pro Gln Thr Ser Glu Ala Phe Ala Ile Thr
1780 1785 1790 1795

Val Arg Asp Val Asn Glu Arg Pro Pro Gln Pro Gln Ala Ser Val Pro
1800 1805 1810

Leu Arg Leu Thr Arg Gly Ser Arg Ala Pro Ile Ser Arg Ala Gln Leu
1815 1820 1825

Ser Val Val Asp Pro Asp Ser Ala Pro Gly Glu Ile Glu Tyr Glu Val
1830 1835 1840

Gln Arg Ala Pro His Asn Gly Phe Leu Ser Leu Val Gly Gly Gly Leu
1845 1850 1855

Gly Pro Val Thr Arg Phe Thr Gln Ala Asp Val Asp Ser Gly Arg Leu
1860 1865 1870 1875

Ala Phe Val Ala Asn Gly Ser Ser Val Ala Gly Ile Phe Gln Leu Ser
1880 1885 1890

Met Ser Asp Gly Ala Ser Pro Pro Leu Pro Met Ser Leu Ala Val Asp
1895 1900 1905

Ile Leu Pro Ser Ala Ile Glu Val Gln Leu Arg Ala Pro Leu Glu Val
1910 1915 1920

Pro Gln Ala Leu Gly Arg Ser Ser Leu Ser Gln Gln Gln Leu Arg Val
1925 1930 1935

Val Ser Asp Arg Glu Glu Pro Glu Ala Ala Tyr Arg Leu Ile Gln Gly
1940 1945 1950 1955

Pro Gln Tyr Gly His Leu Leu Val Gly Gly Arg Pro Thr Ser Ala Phe
1960 1965 1970

Ser Gln Phe Gln Ile Asp Gln Gly Glu Val Val Phe Ala Phe Thr Asn
1975 1980 1985

Phe Ser Ser Ser His Asp His Phe Arg Val Leu Ala Leu Ala Arg Gly
1990 1995 2000

Val Asn Ala Ser Ala Val Val Asn Val Thr Val Arg Ala Leu Leu His
2005 2010 2015

Val Trp Ala Gly Gly Pro Trp Pro Gln Gly Ala Thr Leu Arg Leu Asp
2020 2025 2030 2035

Pro Thr Val Leu Asp Ala Gly Glu Leu Ala Asn Arg Thr Gly Ser Val
2040 2045 2050

Pro Arg Phe Arg Leu Leu Glu Gly Pro Arg His Gly Arg Val Val Arg
2055 2060 2065

Val Pro Arg Ala Arg Thr Glu Pro Gly Gly Ser Gln Leu Val Glu Gln
2070 2075 2080

Phe Thr Gln Gln Asp Leu Glu Asp Gly Arg Leu Gly Leu Glu Val Gly
2085 2090 2095

Arg Pro Glu Gly Arg Ala Pro Gly Pro Ala Gly Asp Ser Leu Thr Leu
2100 2105 2110 2115

Glu Leu Trp Ala Gln Gly Val Pro Pro Ala Val Ala Ser Leu Asp Phe
2120 2125 2130

Ala Thr Glu Pro Tyr Asn Ala Ala Arg Pro Tyr Ser Val Ala Leu Leu
2135 2140 2145

Ser Val Pro Glu Ala Ala Arg Thr Glu Ala Gly Lys Pro Glu Ser Ser
2150 2155 2160

Thr Pro Thr Gly Glu Pro Gly Pro Met Ala Ser Ser Pro Glu Pro Ala
2165 2170 2175

Val Ala Lys Gly Gly Phe Leu Ser Phe Leu Glu Ala Asn Met Phe Ser
2180 2185 2190 2195

Val Ile Ile Pro Met Cys Leu Val Leu Leu Leu Leu Ala Leu Ile Leu
2200 2205 2210

Pro Leu Leu Phe Tyr Leu Arg Lys Arg Asn Lys Thr Gly Lys His Asp
2215 2220 2225

Val Gln Val Leu Thr Ala Lys Pro Arg Asn Gly Leu Ala Gly Asp Thr
2230 2235 2240

Glu Thr Phe Arg Lys Val Glu Pro Gly Gln Ala Ile Pro Leu Thr Ala
2245 2250 2255

Val Pro Gly Gln Gly Pro Pro Pro Gly Gly Gln Pro Asp Pro Glu Leu
2260 2265 2270 2275

Leu Gln Phe Cys Arg Thr Pro Asn Pro Ala Leu Lys Asn Gly Gln Tyr
2280 2285 2290

Trp Val

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..27
(D) OTHER INFORMATION:/product= "Primer P24"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CCTCCAGGTG GTTCTCACCG AAGAAGG

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RA 14"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CCAGACGCCC AACCCGCCAC GATG

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer P23"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCTGTGTGG TGAGTGTAAA CTCC

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..27
(D) OTHER INFORMATION:/product= "Primer RA2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCGCCACGAT GCTTCTCAGC CCGGGAC

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RA12"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGCCTTGTTG GTCAGATCTA CAGC

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer P19"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGGTTCACCA CGTCCAGGAG G

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RA9"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CTCACTGGCT GCCTTCCCTG CCTG

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer RM11"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTCAGTGCTC AGGTACCTCA C

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RA5"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CACGGCGAGC TGGAGCTAGA CATC

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer P9"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GCCACGTCGT CTGTAGTCAG C

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer P10"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GACTCCGCTG ATGGTCTGCA C

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24

(D) OTHER INFORMATION:/product= "Primer P7"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GCAGGTCCTG CTGGGAGAAG GCTG

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer P6"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GTCGAGGTTG GTGTCCAGGT G

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..23
(D) OTHER INFORMATION:/product= "Primer P3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GCTCTTCCAC CATTCTCCAG GAG

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..22
(D) OTHER INFORMATION:/product= "Primer P4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAGGGTCCTG GCTTGAGGTC CG

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..20
(D) OTHER INFORMATION:/product= "Primer BA50"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GTCCTGCTGG GCGTGGAACA

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..18
(D) OTHER INFORMATION:/product= "Primer BA51"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GAACAGCCGG CCTAGCTG

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..33
(D) OTHER INFORMATION:/product= "Primer RMP16"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AGTGAATTCG ATGCAGTCCG GCCGCGGCCC CCC

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RMP19"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CCAGAGAGTG GGGCCCAGAG AAGC

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..25
(D) OTHER INFORMATION:/product= "Primer RMP17"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGAGAAAGCT TCATAATGGC CATAG

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..23
(D) OTHER INFORMATION:/product= "Primer RMP20"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGGCTTCTCT GGGCCCCACT CTC

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer RMP11"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GTCAGTGCTC AGGTACCTCA C

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..25
(D) OTHER INFORMATION:/product= "Primer RMP18"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CTATGGCCAT TATGAAGCTT TCTCC

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..20

(D) OTHER INFORMATION:/product= "Primer RMP9"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GCAGCTGGAT CCAGCCACGG

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer RMP10"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GTGAGGTACC TGAGCACTGA C

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..20
(D) OTHER INFORMATION:/product= "Primer RMP8"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CCGTGGCTGG ATCCAGCTGC

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..21
(D) OTHER INFORMATION:/product= "Primer RMP12"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGCTCCAGCA CCTGCAGGCT G

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..22
(D) OTHER INFORMATION:/product= "Primer RMP14"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GAGGCAGCTG AGATCAGAAG GG

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..24
(D) OTHER INFORMATION:/product= "Primer RMP15"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GACCTTTGTT CTTCCAGAGG TGGC

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..22
(D) OTHER INFORMATION:/product= "Primer RMP3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCAAAGCTTG GGGCACCTCC AG

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..20
(D) OTHER INFORMATION:/product= "Primer RMP4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CCCTAGAGCT CCAGCTGTCC

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..30
(D) OTHER INFORMATION:/product= "Primer RMP1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CTGGAATTCT TAAGCCTGCC TCCACCTTGG

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1..22
(D) OTHER INFORMATION:/product= "Primer RMP2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTGGAGGTGC CCCAAGCTTT GG

Claims

1. Isolated protein designated melanoma-associated chondroitin sulfate proteoglycan (MCSP) and having substantially the amino acid sequence of the mature protein as set forth in SEQ ID NO:2 or an amino acid mutant of said protein excluding the deletional amino acid mutant with the amino acid sequence extending from the amino acid at position 1594 to the amino acid at position 2293 in SEQ ID NO:2, or a derivative of said protein or mutant, particularly such protein, amino acid mutant, or derivative thereof, which is in a suitably immunogenic form.

2. A method for preparing the protein, mutant or derivative according to claim 1 comprising isolation from a natural source, chemical synthesis and/or recombinant DNA technology.

3. A method for the generation of an antibody which specifically binds to MCSP, said method comprising the step of administering to a mammal a protein or mutant, or a derivative thereof, according to claim 1.

4. A pharmaceutical composition comprising a protein or mutant, or a derivative thereof, according to claim 1, and a pharmaceutically acceptable carrier.

5. Nucleic acid comprising an isolated nucleic acid coding for a protein according to claim 1, preferably such nucleic acid which is a DNA.

6. A method for identifying a nucleic acid encoding MCSP, or a novel non-human homologue thereof, comprising contacting a sample comprising candidate DNA or RNA with a nucleic acid comprising at least 14 contiguous bases that are the same as (or complementary to) any 14 or more contiguous bases set forth in SEQ ID NO:1, and identifying nucleic acid(s) which hybridize (s) to said probe.

7. A host cell capable of producing the protein, amino acid mutant, or derivative thereof according to claim 1, and containing a heterologous nucleic acid coding for said protein or mutant, or derivative thereof.

8. The protein, mutant or derivative according to claim 1 for use in the prophylactical or

therapeutical treatment of the human body, particularly for use in the control or (adjuvant) treatment of a MCSP-expressing tumor, such as melanoma, sarcoma and glioblastoma.

9. Use of the protein, mutant or derivative according to claim 1, or a composition comprising said protein, mutant or derivative for the manufacture of a medicament, particularly a medicament suitable for the control or (adjuvant) treatment of a MCSP-expressing tumor, such as melanoma, sarcoma and glioblastoma.

10. A method for identifying a compound which is capable of interacting with the protein, amino acid mutant, or derivative thereof according to claim 1, comprising contacting said protein, mutant or derivative, or a composition of matter comprising said protein, mutant, or derivative, with at least one compound to be tested for its ability to interact with said protein, mutant, or derivative, wherein a change of the biological activity of the protein, mutant, or derivative is indicative of the interaction.

11. Nucleic acid according to claim 5, wherein the isolated nucleic acid is a DNA having substantially the nucleotide sequence set forth in SEQ ID NO:1, or a fragment thereof excluding the fragment with the sequence extending from bp 4867 to bp 7898 in SEQ ID NO:1 and the DNA with the sequence extending from bp 4858 to 5357 in SEQ ID NO:1, respectively..

12. Nucleic acid according to claim 5, which is a hybrid vector.

13. The amino acid mutant according to claim 1, which is a deletional amino acid mutant, or a derivative thereof.

14. A method of vaccinating a human in need thereof comprising the step of administering to said human a protein or mutant, or a derivative thereof, according to claim 1.

15. A test kit for the qualitative or quantitative determination of anti-MCSP antibodies comprising a protein or mutant, or a derivative thereof, according to claim 1.

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/EP 95/03988

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/86 C07K14/82 C07K14/705 A61K38/19
C12Q1/68 C12N5/10 C12Q1/00 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIOLOGY OF PROTEOGLYCANS, 1987, pages 345-366, XP000574176 JOHN R. HARPER ET AL.: "Cell-associated proteoglycans in human malignant melanoma" cited in the application see page 346, last paragraph - page 351, paragraph 2 see page 354, paragraph 3 - page 356, paragraph 1 see page 359, paragraph 1 - page 362, paragraph 2 --- -/--	1,2,8-10

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 1996	09.07.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Montero Lopez, B

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL Database entry HST49467 Accession number T49467; 3 March 1995 HILLIER, L. ET AL.: "WashU-Merck EST Project" XP002006644 see the whole document	5,11
X	EMBL Database entry HS652116 Accession number R53652; 25 May 1995; HILLIER, L. ET AL.: "The WashU-Merck Project" XP002006645 see the whole document	5,11
X	EMBL Database entry HS690124 Accession number R61690; 29 May 1995 HILLIER, L. ET AL.: "The WashU-Merck Project" XP002006646 see the whole document	5,11
A	THE JOURNAL OF CELL BIOLOGY, vol. 114, no. 2, July 1991, pages 359-371, XP000573112 AKIKO NISHIYAMA ET AL.: "The primary structure of NG2, a novel membrane-spanning proteoglycan " cited in the application see abstract see page 359, right-hand column, paragraph 3 see page 361, right-hand column, paragraph 2 - page 365, right-hand column, paragraph 3	1-15
A	EP,A,0 428 485 (NEW YORK MEDICAL COLLEGE) 22 May 1991 cited in the application see the whole document	1-14

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 95/03988

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claim 14 and partially claim 3, are directed to a method of threatment of the human/animal body the search has been carried out and based on the alleged effects of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/EP 95/03988

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-428485	22-05-91	AT-T-	132906	15-01-96
		CA-A-	2029714	15-05-91
		DE-D-	69024758	22-02-96
		ES-T-	2081963	16-03-96
		JP-A-	3206894	10-09-91